

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 23:43:49 ; Search time 914 Seconds
(without alignments)
10183.591 Million cell updates/sec

Title: US-09-931-733-1
Perfect score: 2191
Sequence: 1 tacataatgagagatttta.....tggaggactgaacaagagaaa 2191

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002s.*
7: Geneseq2003as.*
8: Geneseq2003bs.*
9: Geneseq2003cs.*
10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1381.8	63.1	9742	6	ABL70480 Chemical
2	1330.6	60.7	9742	6	ABL70479 Chemical
3	240.2	11.0	2775	4	Aah72785 Human cer
4	240	11.0	2709	6	Abk84059 Human CDN
5	240	11.0	2709	7	Abx10429 DNA encod
6	240	11.0	2709	9	Acf79937 Breast ca
7	239.4	10.9	504	3	Aac00830 Human sec
8	238.4	10.9	2716	6	AAS95004 Human sec
9	230.2	10.5	640	3	Aac00831 Human ova
10	228.9	10.4	2736	6	Abq54618 Human ova
C 11	165.2	7.5	15732	6	Aas45389 Chemical
12	165.2	7.5	15732	6	Abk28234 DNA trans
13	157.4	7.2	15732	4	AAS45388 Chemical
14	157.4	7.2	15732	6	Abx62895 Human act
15	157	7.2	2905	7	Abd47388 Human CDN
16	157	7.2	2906	6	Aac00829 Human sec
17	115.6	5.3	459	3	AbS52426 Human nuc
18	106.4	4.9	1892	6	AAL5558 Human imm
19	106.4	4.9	2817	6	Aak86425 Human imm
C 20	80.2	3.7	768	4	Aak86423 Human imm
C 21	80.2	3.7	768	4	Aak86426 Human imm
C 22	80.2	3.7	768	4	Aak86426 Human imm
C 23	79.4	3.6	175561	7	Aad55694 Human THB

24	77.6	3.5	6073	4	AAK82559 Human imm
25	77.6	3.5	6074	4	AAK82558 Human imm
C 26	77	3.5	7404	4	AAK70264 Human imm
C 27	77	3.5	7409	4	AAK70263 Human imm
C 28	76.8	3.5	1348	4	AAf90649 Human TAN
C 29	76.8	3.5	1348	6	AAD44203 Human TAN
C 30	76.2	3.5	604	5	ABv58969 Human GPC
C 31	75.8	3.5	12932	9	Adc85672 Human GPC
C 32	74.4	3.4	84607	2	AAX90847 Human PAC
C 33	74.2	3.4	110000	6	ABQ74964_1 Continuation (2 of
C 34	74.2	3.4	126512	6	ABN83429 Human am
C 35	73.8	3.4	9377	6	ABA97530 Human am
C 36	73.4	3.4	110000	7	AAI52246_1 Continuation (2 of
C 37	73.2	3.3	488	4	AAI37596 Probe #62
C 38	73.2	3.3	488	4	Aak31720 Human bon
C 39	73.2	3.3	488	4	ABS1404 Human liv
C 40	73.2	3.3	488	6	ABS06476 Human gen
41	73	3.3	300000	9	AdE86352 Human Prp
42	72.8	3.3	136284	6	ABK83575 Human CDN
43	72.8	3.3	249999	7	Abz80229 Human tra
44	72.2	3.3	700	4	Aah93186 Human inf
C 45	71.2	3.2	11481	4	AAK71590 Human imm

ALIGNMENTS

RESULT 1

ABL70480/C standard; DNA; 9742 BP.

ID	ABL70480	AC	ABL70480	DT	01-JUL-2002 (first entry)	DE	Chemically treated cell signalling DNA sequence complementary to#185.	EW	Cell signalling; cytosine methylation; cell signalling disease; cancer; tumour; cytostatic; ds.	OS	Unidentified.	PN	WO200202807-A2.	PD	10-JAN-2002.	PF	29-JUN-2001; 2001WO-EP007471.	PR	30-JUN-2000; 2000DE-01032529.	PS	01-SEP-2000; 2000DE-01043826.	XX	(EPIG-) EPIGENOMICS AG.	XX	Olek A, Piepenbrock C, Berlin K;	XX	WPI; 2002-154758/20.	XX	Nucleic acid, useful for diagnosis and therapy of diseases associated with cell signaling e.g. cancer, comprises chemically modified genomic sequences of genes associated with cell signaling.	XX	Claim 1; SEQ ID NO 370; 24pp + Sequence Listing; English.	XX	The invention relates to a nucleic acid comprising a sequence of at least 18 bases of a segment of chemically pretreated DNA of genes associated with cell signalling. The activity of the modified sequences of the invention may be described as cytostatic. The object of the invention is to provide the chemically modified DNA of genes associated with cell signalling, as well as oligonucleotides and/or PNA-oligomers for detecting cytosine methylations, as well as a method which is particularly suitable for the diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with cell signalling. The chemically modified DNA provided by the invention is useful for diagnosis
----	----------	----	----------	----	---------------------------	----	---	----	---	----	---------------	----	-----------------	----	--------------	----	-------------------------------	----	-------------------------------	----	-------------------------------	----	-------------------------	----	----------------------------------	----	----------------------	----	---	----	---	----	---

CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records AB070311-AB170626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office

XX
Sequence 9742 BP; 2945 A; 88 C; 1819 G; 4890 T; 0 U; 0 Other;
SO

63.1%; Score 1381.8; DB 6; Length 9742; Query Match

Query name	Pred. No. 0;	Gaps
Best Local Similarity	80.8%;	24;
Conservative	0; Mismatches	402;
1702		Indels
		14;

Matches 1787; Conservative

TTTCCACAAATATGCTCGATTAAATAAATTTAGTCACTGTA 60

[illegible]

Db
6482 TACATAATTAATAAAATTTTATTCAGAAATAAGACCTAAATAGATATAAAATTTATATGAGGATTTT

61 ATATCTGCTACATAA-TTATATATATGTATATATATTCGCACA-TTTTTCCTTATCA 110
QY
61 ATATCTGCTACATAA-TTATATATATGTATATATATTCGCACA-TTTTTCCTTATCA 110
QY

D6
6422 ATAATATAACTACATAAATTTATATATATATATATATATATCGCACATTTTACTTTATCA 6366

119 GGAATTTTGTATTTCTACTAATGACTTTTCTGATGATAGCACAATTACAATTGTG 178

6362 AAAATTTTATATATTCTACTAATAACCTTTCTATATAATAACACAATTAACAATTATTA 6303

178 GGCCTGGCCACAAAGCTGGGGGAAATGCACATTTTAAAGCATTTCTTGGTTTCATCTATG 238

6243 CATCTTCTTTAAACATTTCTTTACTTCATCTATA

db 630Z ACCACIACCAGCAGGCTTATTTGGTTGTGCGGCCTATTAGGTG 298

[illegible]

Db
6242 AAAATTCATCTTTATAATAAATTAA/AAAACATTAAGCTAGAAATACCG

QY
299 ACCGCAGTTTGTGCTATGCTAGT-----CCCAATTGGACATTACTGCAATATAAAAAG 333

Db
6183 ACCGCAATTTATATATATATATATATACTAAATCCCAATTAAAAACATTACTACATATAAAATA 6122

353 ATTATAAGTAAAGGCGAGAAAATGTTCTGAAGGGTCATAACTTCCAAGGGTCTCTACC 412

5123 ATTTATAAATAAAAAACA-AAAAATATTCTAAAAAATCATAACTTCCAAAAATTCCTCTACC 6061

DB
0123
412 GGGGACCGTCCAGGTGATCTTCACCTACTGCC 472

QY
413 CTTGACCTTCTGGGTTCCTTAATATCCTTCTCTA CCTACTACC 600

Db
6064 CTTTAUCCCTAAATATAAATCCCCGATTACAGCGTATTTCAC

[illegible]

6004 TTAAACCAACTCAATATATCCAGTAAAAAATAACCTAAAACCACAATAAAAHAIACAC

QY 533 GAACACCCAGAGATGTTTATGGCAT-ATTCAGATCCTCTTTGGCTTTTCTTGACCTTAC 599

5944 GAACACCCGAAAAATATTTATAACCATCATCCAAACCTCTTTACTTTTCTTAACCTTAC 588

592 TCCTTTTGTGATAAAGTTATTCAATATGGTTTCAAATCCAGAATTTCAGCTATTTCACCTTCAT 651

582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333

DD
3003 TCCCTTTTAAATTTT
- - - CCGGGGTCCTTTTGCTCTGTAGATGGCCTTGGACAGT 711

[illegible]

Db
5824 AACCTAATTATAATTAACAAACAACAGCCTTCTTTTCACGCAAAAACAAAAAGCGGGTCTT

QY
712 ACATATTAA CAGTCATAT --- GGGCCCCCTTCCCACTCATGACACATCCCGCTTCAGAACTC

Db 5764 ACATATTACCAATCTATAAAACCCCTTTTCCACTCATAAACCATCTCTACTTACAAATCT 5765

QY 769 GCTGTGGGTAAACACATCATCTGCTAGTTGATGCCATTATTGACATCTCTGGGT 82

5704 ACTACTAAATTAAACACCATCATCTACTAATTAATACCATTTATTTAACATCCTAAAT 56

829 TACAGTACGGCTAAAGGGCACCCAGTGTCTTAAACAGAGAAAATACACTGCTGTGAGCAGAG 88

[illegible]

DB
5644 TACATACAGCCTAATAAAATCACCGGCAATTGGTGTATTTT

[illegible]

QY	1069	CAGCTAGTCTTAACTCCTTCATATAAATTTAAGGAAGGATTTAACATATTCATAATGACT	1128
Db	4339	TAGTTAGTGTAAATTTTTTATATAAATTTAAGNAGGATTTATATATTTATATGATT	4398
QY	1129	TTCTGATCTGGAAAGCCCGCTTGTGCTTATTAATAATTTGATGAATACTTCAGTAG	1188
Db	4399	TTTTGATTTGGAAAGTTTAGTTTGTAGTTATTAATAATTTGATGAATTTTAGGTAG	4458
QY	1189	TATTTAAACAAATAGGATTTCTAGAACATTTCCCAACACAGAGATGACAGTAAAGT	1248
Db	4459	TATTTAAATAAATAGGATTTTATAGTAGATATTTTTTAATTAGGAGATGATGAAAGT	4518
QY	1249	ATTGGTAGCAATAGATCTCTACCAAGATCTCCTGCTGCACTGCCACTCTGTGCACAG	1308
Db	4519	ATTGGTAGTAAATAGATGTTTTTATTAAGATTTTTTTTGTATTTGTTATTTGTGTATAG	4578
QY	1309	GAAAGAGGGGAGAAAGCCATGTTGCTCTCTTATATTTGAGCATCATTTAAGGGAAT	1368
Db	4579	GAAAGAGGGGAGAAAGATTTATGTTGTTTTTATATTTAGTATTTATTTAAGGGAAT	4638
QY	1369	ATAAGTCTTAACTCATGATGATGCTGCTGGGTAAAGAAAGTTCTTAATGTTGGAAT	1428
Db	4639	ATAAGTTTTTAAATGATTTATATGATGTTTGTGGGTAAGAAAGTTTTTAATGTTGGAAT	4698
QY	1429	CCTTTCAGCTATCGAATTTTGTAG-AAAAAACTTATTTGTTGTTTATACATTTATGA	1487
Db	4699	TTTTTTAGTTATCGAATTTTGTAGAAATAAATAATTTATTTGTTTATATATATTTATGA	4758
QY	1488	GATCAAGTGTAAATTCGTTTACATGATATGATGATGATGATGATGATGATGATGAT	1547
Db	4759	GATATAGTGTAAATTCGTTTATATGATATATGATGATGATGATGATGATGATGAT	4818
QY	1548	TGTCGATCATCCGAAATTTAACTGTACCATTAAGTAAATTTCTATCCCCCAT	1607
Db	4819	TGTCGATCATCCGAAATTTAACTGTACCATTAAGTAAATTTCTATCCCCCAT	4878
QY	1608	TCCCTCACCACATGCTCCAGCCCTTGCAGGAACTGTTCAATTTCTGACCTAACAGA	1667
Db	4879	TTTTTTTATATATGTTTATGTTTGTAGGAAATGTTTATTTTGTATTAATAGA	4938
QY	1668	AACGAAAGCTAAAAACATGTTGGAGGAGTCTCCACATGTTTCCCTACTCCATTTCTC	1727
Db	4939	AACGAAAGCTAAAAATTTATGTTGGAGGAGTTTATATGTTTATTTATTTATTTT	4998
QY	1728	TGGGG-ATAGCAGATAGGACCAAGCCAGCAGTACTAGTCACTAACTAGTCACTCAACA	1786
Db	4999	TGGGGTAAATAGATAGGATAGGATAGTATGATTTAGTTATTAATTAAGTATTAATTA	5058
QY	1787	AGGCCCTTTTTCCTTTGTTATCTTTTCAGATATCTCATTTCTTCTAGGATTA	1846
Db	5059	AGGTTTTTTTTTTTGTATTTTGTAGATATTTTATTTTATGCTTTTGTAGGATTA	5118
QY	1847	CAACATCCCGGTTCCGTTCTGGAACTTTTACTGATTTATCTCCCCCTCACACAAAT	1906
Db	5119	TAATATTTTTCGCTTTCGTTTGTGGAAATTTTATGATTTATTTTATTTATTAAT	5178
QY	1907	RAGCATTTGATCTGCTGATTTCTGAGATCTCAAGATCTGAGTACTGTTGTGAAATAATTT	1965
Db	5179	RAGTATTTGATTTTGTATTTTGTAGATTTTAAAGATTTTGAATTTGTTGTTGTTT	5238
QY	1966	---CCAGGTGAGTACTGTTCTGATTTTGAATATGATCTGTTCTCTCTGAG-TC	2021
Db	5239	TAGTGAAGTGAATTTGTTTTCATTTTGTAAATATGATTTTGTGTTTTTTTTTTTGAAGTT	5298
QY	2022	CCAGATCAACAGGGGACATAGTATT-CTTATTCAGGTCATGAGTATGAGTGGAGTAG	2080
Db	5299	TTAGAAATTAAGGGGATTAATAGTATGTTTATTTAGGGTTATGAGTATGAGGAGTAG	5358
QY	2081	GGCTGAGTATTCAGAAAGTG-AAACTGAGTGTGCTGATATGAATCTTCAATTTACTTAG	2139
Db	5359	GGTTGAGTATTTAGAAAGTGAAATTTGAGTTGTTGATATGAATTTTATTTATTTATAG	5418
QY	2140	GAAGATACAGGATCTCTTA-TTCACCACTAGGACTGAAACAGAGAAA	2191
Db	5419	GAAGATAATAGGATTTTTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	5471
RESULT 3			
AAH72785			
ID	AAH72785	standard; cDNA; 2775 BP.	
XX	AAH72785;		
AC	AAH72785;		
DT	19-SEP-2001	(first entry)	
XX	Human cervical cancer marker nucleic acid 4059.		
DE	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.		
XX	Homo sapiens.		
OS	WO200142467-A2.		
XX	14-JUN-2001.		
PD	08-DEC-2000; 2000WO-US033312.		
PF	08-DEC-1999; 99US-0169681P.		
XX	21-DEC-1999; 99US-01711350P.		
PR	14-MAR-2000; 2000US-0189315P.		
PR	12-MAY-2000; 2000US-0203791P.		
PR	09-JUN-2000; 2000US-0210600P.		
PR	21-JUL-2000; 2000US-0220114P.		
XX	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.		
PA	Schlegel R, Deeds J, Berger A, Zhao X;		
XX	WPI; 2001-375006/39.		
PI	New isolated nucleic acid for diagnosing and treating cervical cancer and		
PT	for assessing and detecting compounds for treating the cancer.		
XX	Claim 1; Page 850-851; 1051pp; English.		
PS	The invention relates to novel genes (AAH68727-AAH73383) associated with		
XX	cervical cancer with cytostatic activity. The nucleic acids and encoded		
CC	polypeptides are useful: to assess if a patient is afflicted with		
CC	cervical cancer or has a pre-malignant condition; to monitor the		
CC	progression of cervical cancer or a premalignant condition in a patient;		
CC	and to select and/or assess the efficacy of a compound or therapy for		
CC	inhibiting cervical cancer in a patient. The nucleic acids may also be		
CC	useful for gene therapy		
XX	Sequence 2775 BP; 971 A; 559 C; 566 G; 664 T; 0 U; 15 Other;		
SQ	Query Match 11.0%; Score 240.2; DB 4; Length 2775;		
	Best Local Similarity 93.3%; Pred. No. 1.9e-50;		
	Matches 251; Conservative 0; Mismatches 18; Indels 0; Gaps 0;		
QY	1701	CCACATTTGTTTCTTACTCCATTTTCTCTGGGGAATAGCAGAAATAGGACAGCAGCACT	1760
Db	10	CCGCGCTCGACTACGGCTCGGAGAACCCCGGGAATAGCAGAAATAGGACAGCAGCACT	69
QY	1761	ACTCAGCTAACTAAGTGAATCAACCAAGGCCCTTTTCTTGTATCTTGTGAGATCTT	1820
Db	70	ACTCAGCTAACTAAGTGAATCAACCAAGGCCCTTTTCTTGTATCTTGTGAGATCTT	129
QY	1821	CATTTTCTTGTGAGTCTTGTGAGTCAACATCTCTCGGTTCCGTTCTGGAACTTTAC	1880
Db	130	CATTTTCTTGTGAGTCTTGTGAGTCAACATCTCTCGGTTCCGTTCTGGAACTTTAC	189
QY	1881	TGATTTATCTCCCTCAGCAAAATAGCAATTTGATTTCTGATTTCTGAGAACTCTCAG	1940
Db	190	TGATTTATCTCCCTCAGCAAAATAGCAATTTGATTTCTGATTTCTGAGAACTCTCAG	249


```

1941 ATCTGGACTACTGTGAAAAAATTCAG 1969
      |||||||
250 ATCTGGACTACTGTGAAAAAATTCAG 278

RESULT 4
ABK84059
ID ABK84059 standard; cDNA; 2709 BP.
XX
AC ABK84059;
XX
XX 14-AUG-2002 (first entry)
XX
XX Human cDNA differentially expressed in granulocytic cells #630.
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
XX viral infection; parasitic infection; protozoal infection;
XX fungal infection; sterile inflammatory disease; psoriasis;
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;
XX adult respiratory distress syndrome; inflammatory bowel disease;
XX Crohn's disease; ulcerative colitis; periodontal disease;
XX granulocyte activation; chronic inflammation; allergy.
XX
XX Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US030821.
XX
XX 03-OCT-2000; 2000US-0237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
XX genes associated with granulocyte activation, which serves as diagnostic
XX markers that is useful for monitoring disease states and drug toxicity.
XX
XX Claim 1; SEQ ID NO 630; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing the
XX expression level to an expression level in an unactivated GC, where
XX differential expression of Gs is indicative of GCA. Also included are
XX modulating (M2) GA by contacting GC with an agent that alters the
XX expression of at least one gene in Gs; (2) screening (M3) for an agent
XX capable of modulating GCA or an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease using the gene expression
XX profile; (3) detecting (M4) an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease, by detecting the level of
XX expression in a sample of the tissue of gene(s) from Gs, where the level
XX of expression of the gene is indicative of inflammation; (4) treating
XX (M5) an inflammation (especially chronic) or in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease, by contacting a tissue having inflammation with an
XX agent that modulates the expression of gene(s) from Gs in the tissue. M1
XX is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
XX for screening an agent capable of modulating GCA preferably in an
XX inflammation in a tissue; M4 is useful for detecting an inflammation
XX (especially chronic) in a tissue, an allergic response in a subject,
XX exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
XX psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
XX cardiac reperfusion injury, renal reperfusion injury, ARDS, adult

```

```

CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease, also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 2709 BP; 960 A; 542 C; 548 G; 659 T; 0 U; 0 Other;
XX
Query Match 11.0%; Score 240; DB 6; Length 2709;
Best Local Similarity 100.0%; Pred. No. 2.1e-50;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1730 GGGAAATAGCAGATAGGAGCAGCAGCAGCAGTACTAGTACGACTCAACCAAGG 1789
Db 1 GGGAAATAGCAGATAGGAGCAGCAGCAGCAGTACTAGTACGACTCAACCAAGG 60
QY 1790 CCTTTTTCCTTGTATCTTTTCGAGATACATTTCTTAGCGTTTCTGGAGATTACAA 1849
Db 61 CCTTTTTCCTTGTATCTTTTCGAGATACATTTCTTAGCGTTTCTGGAGATTACAA 120
QY 1850 CATCTCGCGTTCCGTTTCTCGGAACTTTACTGATTTATCTCCCTCCACAAATTAAG 1909
Db 121 CATCTCGCGTTCCGTTTCTCGGAACTTTACTGATTTATCTCCCTCCACAAATTAAG 180
QY 1910 CATTGATTCCTGCATTTCTGGAATCTCAAGATCTGGACTACTGTGAAAAAATTCAG 1969
Db 181 CATTGATTCCTGCATTTCTGGAATCTCAAGATCTGGACTACTGTGAAAAAATTCAG 240

RESULT 5
ABX10429
ID ABX10429 standard; DNA; 2709 BP.
XX
AC ABX10429;
XX
XX 28-JAN-2003 (first entry)
XX
XX DNA encoding protein differentially regulated in prostate cancer #98.
XX
XX Prostate cancer; gene expression; differential regulation;
XX molecular marker; drug target; cancer detection; cancer diagnosis;
XX cancer staging; cancer grading; cancer assessing; cancer monitoring;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO200281638-A2.
XX
XX 17-OCT-2002.
XX
XX 08-APR-2002; 2002WO-US010824.
XX
XX 06-APR-2001; 2001US-0281731P.
XX
XX 06-APR-2001; 2001US-0281732P.
XX
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
XX Sun Z, Jay G;
XX WPI; 2003-058520/05.
XX
XX Novel genes which are differentially regulated in prostate cancer, useful
XX for diagnosing prostate cancer in prostate tissue sample and assessing
XX therapeutic or preventive intervention in prostate cancer patients.
XX
XX Claim 1; Page 191-192; 416pp; English.
XX
XX The invention describes genes (I) which are differentially regulated in
XX prostate cancer. (I) is useful for diagnosing a prostate cancer in a

```

sample comprising prostate tissue, which involves determining the number of target genes which are differentially-regulated in the sample, where the number is indicative of the probability that the sample comprises prostate cancer. (I) Is useful for assessing a therapeutic or preventive intervention in a subject having a prostate cancer, which involves determining the expression levels in a sample comprising prostate tissue of target genes which are differentially-regulated in prostate cancer. Preferably, the expression levels of at least 10 genes are determined. (I) is also useful for identifying agents that modulate a biological activity of a polypeptide differentially-regulated in prostate cancer cells, which involves contacting a polypeptide differentially-regulated in prostate cancer cells with a test agent under conditions effective for the test agent to modulate a biological activity of the polypeptide, and determining whether the test agent modulates the biological activity. (I) is useful as molecular markers, as drug targets, and for detecting, diagnosing, staging, grading, assessing, monitoring, prognosticating, preventing or treating, determining predisposition to diseases and conditions especially relating to prostate cancer. (I) and its expression products are used in the diagnostic test to assay for presence of cancer e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in blood etc. (I) is useful for assessing cancer e.g., to determine the type of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways physiologically relevant to prostate cancer permits the definition of functional and disease pathways and the delineation of targets in these pathways which are useful in diagnostic, therapeutic, and clinical applications. This sequence encodes a protein differentially regulated in prostate cancer

Sequence 2709 BP; 960 A; 542 C; 548 G; 659 T; 0 U; 0 Other;

Query Match 11.0%; Score 240; DB 7; Length 2709;
Best Local Similarity 100.0%; Pred. No. 2.1e-50;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1730 GGGAAATAGCAGATAGGAGCAAGCCAGCAGTACGCTAAGTACTAAGTACTCAACCAAGG 1789
1 GGGAAATAGCAGATAGGAGCAAGCCAGCAGTACGCTAAGTACTAAGTACTCAACCAAGG 60

1790 CCTTTTTCCTTGTTATCTTTGCGAGATCTTCATTTTCTTAGCGTTTCTGGAGATTACAA 1849
61 CCTTTTTCCTTGTTATCTTTGCGAGATCTTCATTTTCTTAGCGTTTCTGGAGATTACAA 120

1850 CATCTCGGGTTCGGTTTCTGGGAACCTTACTGATTTATCTCCCTCCACACAAATAAG 1909
121 CATCTCGGGTTCGGTTTCTGGGAACCTTACTGATTTATCTCCCTCCACACAAATAAG 180

1910 CATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 1969
181 CATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 240

RESULT 6
ACF79937
ID ACF79937 standard; cDNA; 2709 BP.

XX ACF79937;

DT 15-JAN-2004 (first entry)

DE Breast cancer specific gene IFI16 under-expressed in breast cancer.

XX Breast cancer; marker; human; diagnosis; cytostatic; biochip; vaccine;

KW IFI16; gene; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FS 265..2454

FT /*tag= a
FT /product= "IFI16"

XX WO2003073911-A2.

XX 12-SEP-2003.

XX 27-FEB-2003; 2003WO-US005984.

XX 28-FEB-2002; 2002US-0359999P.

XX (GEOU) UNIV GEORGETOWN.

XX Su YA, Yang J;

XX WPI; 2003-721995/68.

XX P-PSDB; ABW78957.

DR Detecting breast cancer in a subject comprises contacting a biological sample with an agent that binds to a polynucleotide or polypeptide of a breast-cancer specific gene (BCSG).

XX Claim 6; Page 95-96; 143pp; English.

XX The present sequence is that of a breast cancer specific gene (BCSG) identified by microarray gene expression analysis as being under-expressed in breast cancers in comparison to healthy tissue. The gene was identified as IFI16. It is 1 of 19 (see ACF79921-39) BCSGs of the invention that are differentially expressed in breast cancer cell lines and breast cancer tissue samples as compared to control cell lines and normal tissue samples. The polynucleotides transcribed from the BCSGs and the polypeptides encoded by the BCSGs are designated as breast cancer specific markers (BCSM). The invention provides a method for detection of breast cancer by measuring expression levels of BCSGs, and in particular the levels of polynucleotides transcribed from and polypeptides encoded by the BCSGs. A pharmaceutical composition for the treatment of breast cancer comprises a BCSM, an antibody directed against a BCSM, a vaccine generated using a BCSM, or an agent that modulates an expression level of a BCSG or an activity of a BCSM. A biochip for diagnosing breast cancer or screening agents that inhibit breast cancer comprises a BCSG or BCSM

Sequence 2709 BP; 960 A; 542 C; 548 G; 659 T; 0 U; 0 Other;

Query Match 11.0%; Score 240; DB 9; Length 2709;
Best Local Similarity 100.0%; Pred. No. 2.1e-50;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1730 GGGAAATAGCAGATAGGAGCAAGCCAGCAGTACGCTAAGTACTAAGTACTCAACCAAGG 1789
Db 1 GGGAAATAGCAGATAGGAGCAAGCCAGCAGTACGCTAAGTACTAAGTACTCAACCAAGG 60

QY 1790 CCTTTTTCCTTGTTATCTTTGCGAGATCTTCATTTTCTTAGCGTTTCTGGAGATTACAA 1849

Db 61 CCTTTTTCCTTGTTATCTTTGCGAGATCTTCATTTTCTTAGCGTTTCTGGAGATTACAA 120

QY 1850 CATCTCGGGTTCGGTTTCTGGGAACCTTACTGATTTATCTCCCTCCACACAAATAAG 1909

Db 121 CATCTCGGGTTCGGTTTCTGGGAACCTTACTGATTTATCTCCCTCCACACAAATAAG 180

QY 1910 CATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 1969

Db 181 CATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 240

RESULT 7
AAC00830
ID AAC00830 standard; cDNA; 504 BP.

XX AAC00830;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 828.

DE

XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EPI033401-A2.
XX PD 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GSEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX P-PSDB; AAG00824.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 1; SEQ ID NO 828; 71pp + Sequence Listing; English.
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or poly(A)⁺ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors
XX
SQ Sequence 504 BP; 163 A; 90 C; 97 G; 154 T; 0 U; 0 Other;
Query Match 10.9%; Score 239.4; DB 3; Length 504;
Best Local Similarity 99.6%; Pred. No. 1.5e-50;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1729 GGGCAATAGCAGATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAG 1788
Db 11 GGGCAATAGCAGATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAG 70
QY 1789 GCGTTTTTCCCTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 1848
Db 71 GCGTTTTTCCCTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 130
QY 1849 ACATCTCGGTTCCGTTTCTGGGAACCTTACTGATTATCTCCCGCTCACACAAATAA 1908
Db 131 ACATCTCGGTTCCGTTTCTGGGAACCTTACTGATTATCTCCCGCTCACACAAATAA 190
QY 1909 GCATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTCTGTGAAAAAATTTCCA 1968
Db 191 GCATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTCTGTGAAAAAATTTCCA 250
QY 1969 G 1969
Db 251 G 251
RESULT 8
AAS95004
ID AAS95004 standard; DNA; 2716 BP.
XX

AC AAS95004;
XX 14-FEB-2002 (first entry)
XX Human DNA sequence #259 expressed during foam cell differentiation.
XX Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX Homo sapiens.
XX WO200177389-A2.
XX 18-OCT-2001.
XX 04-APR-2001; 2001WO-US011128.
XX 05-APR-2000; 2000US-0195106P.
XX (INCY-) INCYTE GENOMICS INC.
XX Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
PI Tai J;
XX WPI; 2002-010925/01.
XX Composition useful for diagnosis of conditions, disorders or diseases
PT associated with atherosclerosis, comprises several polynucleotides that
PT are differentially expressed in foam cell development.
XX Claim 1; Page 300-301; 315pp; English.
XX The present invention relates to the isolation of human polynucleotide
CC sequences that are differentially expressed during foam cell
CC differentiation. The polynucleotide sequences of the invention or a
CC composition comprising these polynucleotides are useful as a high
CC throughput method for detecting altered expression of one or more
CC polynucleotides in a sample. The polynucleotides can be used in the
CC diagnosis of disorders associated with foam cell development such as
CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC coronary artery disease. The polynucleotide sequences can also be used as
CC PCR primers and probes. The polynucleotides of the invention are also
CC useful in gene therapy. AAS94746-AAS95021 represent the human
CC polynucleotide sequences of the invention which are differentially
CC expressed during foam cell differentiation
XX
SQ Sequence 2716 BP; 964 A; 537 C; 548 G; 666 T; 0 U; 1 Other;
Query Match 10.9%; Score 238.4; DB 6; Length 2716;
Best Local Similarity 99.6%; Pred. No. 5.4e-50;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1730 GGGCAATAGCAGATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAGG 1789
Db 1 GGGCAATAGCAGATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAGG 60
QY 1790 CCGTTTTTCCCTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 1849
Db 61 CCGTTTTTCCCTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 120
QY 1850 CATCTCGGTTCCGTTTCTGGGAACCTTACTGATTATCTCCCGCTCACAAATAAG 1909
Db 121 CATCTCGGTTCCGTTTCTGGGAACCTTACTGATTATCTCCCGCTCACAAATAAG 180
QY 1910 CATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTCTGTGAAAAAATTTCCAG 1969
Db 181 CATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTCTGTGAAAAAATTTCCAG 240
RESULT 9
AAC00831
ID AAC00831 standard; cDNA; 640 BP.
XX

SQ Sequence 2736 BP; 989 A; 528 C; 543 G; 672 T; 0 U; 4 Other;

Query Match 10.4%; Score 228.8; DB 6; Length 2736;
 Best Local Similarity 93.4%; Pred. No. 1.5e-47;
 Matches 239; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 805 GCATTTATTGATCTCTGGGTTACAGCAGCTAAAGGCAACCCAGTGCTTAAACAG 864
 DB 1 GCATTTATTGATCTCTGGGTTACAGCAGCTAAAGGCAACCCAGTGCTTAAACAG 60

QY 865 AGAAATACATCTGTGAGCAGAGGATTAAACAGACTCTCTCACTACCCGGAAGTTCACT 924
 DB 61 AGAAATACATCTGTGAGCAGAGGATTAAACAGACTCTCTCACTACCCGGAAGTTCACT 120

QY 925 GCATTTTCAGGATTTTCAAGTCTTCGATGCTGTCTATGTGAGTCCACTATCCAGTT 984
 DB 121 GCATTTTCAGGATTTTCAAGGCTTCGATGCTGTCTATGTGAGTCCACTATCCAGTT 180

QY 985 TGTATCCAGGAATTAATGCAGAAATCTGAGCAATTCGTGAATCTAATCAATGAGGTAA 1044
 DB 181 TGTATCCAGGAATTAATGCAGAAATCTGAGCAATTCGTGAATCTAATCAATGAGGTAA 240

QY 1045 GTAACCTAGTCATGGT 1060
 DB 241 TTCAATTTCTTAGCGT 256

RESULT 11
 AAS45389/c
 ID AAS45389 standard; DNA; 15732 BP.
 XX
 AC AAS45389;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Chemically pretreated complementary DNA associated with cell cycle #47.
 XX
 KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KW PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200168911-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP002945.
 XX
 PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019038.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602751/68.
 XX
 PT Designing primers and probes for analyzing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle.
 XX
 PS Claim 1; SEQ ID NO 94; 28pp; English.
 XX
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the

CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers
 XX

SQ Sequence 15732 BP; 5557 A; 70 C; 2865 G; 7240 T; 0 U; 0 Other;

Query Match 7.5%; Score 165.2; DB 4; Length 15732;
 Best Local Similarity 56.8%; Pred. No. 4.3e-31;
 Matches 586; Conservative 0; Mismatches 388; Indels 58; Gaps 13;

QY 442 AGTGCTTAATATCTCTTCTTCACTTCTGCTTACCTTACCTTAAAGCCAACTCAATGTATCCAGTAA 501
 DB 11919 AATTTCTTATATCTCTTCTTACCTTACCTTAAAGCCAACTCAATGTATCCAGTAA 11860

QY 502 GAGTACCTTAAAGCCAGTAAAGATTATCAGAACACCCAGAGATGTTTATGCCATATT 561
 DB 11859 AAAAAAATCAA-----ACATTAACACCCACATATTAATTAATCAT 11814

QY 562 CCAGATCTCTTGTCTTTTCTTGACTTACTCTTTTGTATAAGTATTATCATATGTT 621
 DB 11813 TCTAAATCTCTTAATTTTCTTAAATTTTACCTTACTTAAATAAATTTAAACA---ATT 11757

QY 622 TCAATCCAGATTCAGCTATTCATCTCATGCCCTAGTTGGATTAAACATCAGCTT 681
 DB 11756 CCAATCCAAACTTTA---ACTCTCTATATCTACTAATTTAAATTAACCAATCACTT 11700

QY 682 TCTTTTGTCTGTAGATGGCTTGGACAGT-----ACATATTAAACAGTCTATGGGCCCTT 736
 DB 11699 TATTTCTCTATAAATAACCTTAATCATTTTCCCTTATTAATCTATTTAAATTCCTTT 11640

QY 737 TCCATCTATGACCACTCTCTGCTTGCAGATCTGTCTGGGTAAACACCAATCACTGC 796
 DB 11639 TCCATCATCAACCACTTTTACTTACAAATCAACTAATTAATAACCAACACTATCTAC 11580

QY 797 TAGTTGATGCCATTTATTTGATCATCTCTGGGTTCAGACACGCTAAAGGCCACCACTGTC 856
 DB 11579 TAACTAATACCACTTCAATTTAACTTAACTTAACTTAACTTAACTTAACTTAACTT 11520

QY 857 CTAAACAGAGAAAAATACACTGCTGAGCAGAGGATTAAACAGACTCTCTCACTACCGGA 916
 DB 11519 CTTCACAA--AAAAAATAATTACAAATAAATAAATTTATCAACCCCTTATCATTTCAA 11461

QY 917 AGTTCAAGTGCAATTTCCAGGATTTTCAAGTCTTCAATCTTCAATCTTCAATCTTCAAT 976
 DB 11460 ATTTCAATAC-TATTTCCAAAATTTTCAAAATTTCAATCATATAATAAACCACAA----- 11410

QY 977 TCCAGTTTGTGTTTCCAGGAAATATGCAGAAATCTGAGCATTCGTGAATCTAATCAAT 1036
 DB 11409 -----CTATCCATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATC 11361

QY 1037 GAGGTAAAGTAACTAGTCTATGG-----TCAAGAGGACAGCTAGTGTAACTCTCTTATAT 1093
 DB 11360 AAAAAAACAACCTAATTTATCTCAAAAAAACAACCTAATATAAATCTTTTCAAT 11301

QY 1094 AATTTAAGGAAGGATTTACATATTCATATTAATGATTTCT--GATCTGGAAGCCCGCT 1151
 DB 11300 -TATTTAAATAAAAAATTTACATATTCATATAAATCTTCTACCTCTATCAAACTCCAAATA 11242

QY 1152 TGTCAAGTTTAAATAATTTGGATGAATCTTCAAGTAGTATTTAACAAATAGCATTTCA 1211
 DB 11241 TATCAATTTAATAACTAAATAAATA-TATAAATATATTTTAAACAAA-AACAATTTCA 11184

QY 1212 TAGAAGACATTTCCCAACAGGAGATGCAGTAAAGTATTGGTAGCAATAGATGCTCTTA 1271

Db 11183 CAATAAACCCACCCTATCCAAAAATAAATCAAAATTTCTAATAACAATAAATCTCTA 11124
 QY 1272 CCAAGATCTCTCTGCTGCACTGCACTCTGTGCACAGGAAGGAGGAGGAAGCCAT 1331
 Db 11123 CCAAAATCTCTACTCTCTACCTCTATAAACAACAAAAAATAAATTTAATATT 11064
 QY 1332 GTTGCTCTCTTATTTG-----AGCATATTTAAGGGAATATAAGTCTCTAATGACTCA 1387
 Db 11063 TACCTTTCCCAATTTATTAATAATCTTATCAACAGCACTACCGATTTCTAATAACCTA 11004
 QY 1388 CAGTGAAGTCTGCTGGTGAAGTCTTAATCTTGAATCTCTCTTCACTATCGAATTT 1447
 Db 11003 TATATAATTAATCTTATAAAAAAATTTCTAAATTTAAACTCTCTTCAACAAATATTAT 10944
 QY 1448 TGTAAAGAAAAA 1459
 Db 10943 AATCAAAAAA 10932

RESULT 12
 ID ABK28234/c
 ABK28234 standard; DNA; 15732 BP.

AC ABK28234;

DT 23-APR-2002 (first entry)

XX DNA transcription associated complementary genomic DNA #54.

XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.

XX Unidentified.

XX WO200192565-A2.

XX 06-DEC-2001.

XX 06-APR-2001; 2001WO-EP003973.

XX 06-APR-2000; 2000DE-01019058.

XX 07-APR-2000; 2000DE-01019173.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-090046/12.

XX New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological disorders,
 PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
 PT cancer.

XX Claim 1; SEQ ID NO 108; 32pp; English.

XX The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer, that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)

CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification but
 CC was obtained in electronic format directly from the European Patent
 CC Office
 XX
 SQ Sequence 15732 BP; 5557 A; 70 C; 2865 G; 7240 T; 0 U; 0 Other;

Query Match 7.5%; Score 165.2; DB 6; Length 15732;
 Best Local Similarity 56.8%; Pred. No. 4.3e-31;
 Matches 586; Conservative 0; Mismatches 388; Indels 58; Gaps 13;

QY 442 AGTGCCCTTAATATCCTTCTCTACCTACTGCTTAAGGCCAACTCAATGTATCCACATAAA 501
 Db 11919 AATTTCCTTATCTCTCTTACCTTTACCTTAAACCTACTCGAAACATCAATATAA 11860
 QY 502 GAGTACCCCTAAAGCCAGTAAAGATTATCAGAACACCCAGAGATGTTTATGGCCATATT 561
 Db 11859 AAAAAAATAAA-----ACAAATAACACCCACATATTAATAATCAATCAT 11814
 QY 562 CCAGATCCCTTGTCTTCTTCTGCTACTCTCTTTTGTAAAGTATTTCATATGGTT 621
 Db 11813 TCTAAATCCTCTTAATTTTCTTAATTTACCTACTTTTAATAAATTTAAACA---ATTT 11757
 QY 622 TCAAAATCCAGAAATTCAGCTATTACCTTCATGCGCCCTAGTTTGGATTAAACATCAGCTT 681
 Db 11756 CCAATCCAACTTTA--ACITCTCTATACTACTAATTTAAATTTAAACAACTCACTT 11700
 QY 682 TCTTTTGTCTGTAGATGGCTTTGGACAGT-----ACATATTACAGTCTATGGGCCCTT 736
 Db 11699 TATTTCTCTATATAATAACCTTAAATCATTTTCCCTTATTAATATTAAAAATCCTTT 11640
 QY 737 TCCACTCATGACACACATCCTGCTTGAGATCTGCTGCTGTTAAACACACATCTCTCTG 796
 Db 11639 TCCATCATACACACATTTTACTTTACAAATCAACTAATTAATTAACACACACTCTCTAC 11580
 QY 797 TAGTTGATGCCATTTTATTTGACATCTCTGGGTTTACAGCAGCTTAAAGGCCACCTCTGTC 856
 Db 11579 TAACTAATACCATTTCAATTTAAACAATTTCAAAATTTAATACTTTTATAATACACCACTATT 11520
 QY 857 CTTTACAGAGAAAAATPACACTGCTGAGCAGAGAGATTAAACAGACTCTCTCACACTACGGGA 916
 Db 11519 CTTTCAAAA-AAAAAATAATTAACAATACAAAAAATTTATCAAAACCCCTATCATTTCCAAA 11461
 QY 917 AGTTCAGTGCAATTTTCCAGGATTTTCAAGTCTTCGATGCTGTCATGTTGAGAGTCCACTA 976
 Db 11460 ATTTCAATAC-TATTCAAAATATTCAAAATTTCAATCATATATATAAACCAA----- 11410
 QY 977 TCCACCGTTTGTATCCAGGAATPACAGAAAAATCGAGCAATTCGTGAATCTTAATCAAT 1036
 Db 11409 -----CTATCCATAAATAATACAAAAAATAAATAATATAAATTTATTAATC 11361
 QY 1037 GAGTAAAGTAACCTAGTCTATGG---TCAAGAGGACAGCTAGTGTGTTAAACTCTTTCATAT 1093
 Db 11360 AAAATAAAACAACTCAATTTATATATCTCAAAAAAACAATTAATAAATACTCTTTTCAAT 11301
 QY 1094 AAATTTAAAGGAGGATTTTACATATTCATATGACTTTCT--GATCTGGAAGGCCAGGCT 1151
 Db 11300 -TATTTAAATAAATAATTTACATTTTCTTAATAATTTCTTACCTCTATCAAACTCCAAATA 11242
 QY 1152 TGTCAAGTTTAAATAATGGAATGAACTCTTCAAGGTAGTATTTAAACAAATAGCGATTCTCA 1211
 Db 11241 TATCAATTTAATAACTAAATAAATA-TATAAATACTATTTAAACAAATAA-AACAATTTCA 11184

Search completed: March 28, 2004, 01:38:52
Job time : 922 secs

```
RESULT 15
ABX62895
ID   ABX62895 standard; cDNA; 2905 BP.
XX
AC   ABX62895;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human activated T cell cDNA #11.
XX
KW   T cell; gene; es; differential expression; T cell activation;
KW   antiallergic; cytostatic; immunosuppressive; antimicrobial; gene therapy;
KW   allergy; cancer; graft versus host disease; infection;
KW   autoimmune disorder.
XX
OS   Homo sapiens.
XX
PN   US2002137077-A1.
XX
PD   26-SEP-2002.
XX
PF   25-OCT-2001; 2001US-00002600.
XX
PR   25-OCT-2000; 2000US-0243521P.
XX
PA   (HOPK/) HOPKINS C M.
PA   (PETE/) PETERSON D P.
PA   (COCK/) COCKS B G.
PA   (HAWK/) HAWKINS P R.
XX
PI   Hopkins CM, Peterson DP, Cocks BG, Hawkins PR;
XX
DR   WPI; 2003-102381/09.
XX
PT   New combination comprising several cDNAs that are differentially
PT   expressed in activated T cells, useful for diagnosing, treating, staging
PT   or monitoring treatment for allergy, cancer, infectious and/or autoimmune
PT   disorders.
XX
PS   Claim 1; Page; 180pp; English.
XX
CC   This invention relates to the sequences of several cDNAs that are
CC   differentially expressed in activated T cells. The sequences of the
CC   invention may have antiallergic, cytostatic, immunosuppressive and
CC   antimicrobial activity and may be used in gene therapy. The invention
CC   also comprises a method for screening samples for differentially
CC   expressed genes and a method for detecting these cDNAs by hybridisation.
CC   The methods and compositions of the present invention are useful for
CC   diagnosing, treating, staging or monitoring treatment for allergy,
CC   cancer, chronic graft versus host disease, infectious and/or autoimmune
CC   disorders. The present sequence represents a cDNA of the invention that
CC   is differentially expressed in activated T cells
XX
SQ   Sequence 2905 BP; 1030 A; 593 C; 577 G; 705 T; 0 U; 0 Other;
Query Match          7.2%; Score 157; DB 7; Length 2905;
Best Local Similarity 100.0%; Pred. No. 2.7e-29;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1813 AGATACCTCATTCTTTAGCGTTTCTCGAGATTACACATCTCGCGTTCCGTTTCTGGG 1872
Db 107 AGATACCTCATTCTTTAGCGTTTCTCGAGATTACACATCTCGCGTTCCGTTTCTGGG 166
QY 1873 AACCTTACTGATTTATCTCCCCCTCACACAAATAGCATTGATCTCTGATTTCTGAAG 1932
Db 167 AACCTTACTGATTTATCTCCCCCTCACACAAATAGCATTGATCTCTGATTTCTGAAG 226
QY 1933 ATCTCAGATCTGGACTACTGTTGAAAAAATTTCCAG 1969
Db 227 ATCTCAGATCTGGACTACTGTTGAAAAAATTTCCAG 263
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 01:12:30 ; Search time 184 Seconds
(without alignments)
6608.138 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 2191

Sequence: 1 tacataattgagagatttta.....tgaggactgaacagagaaa 2191

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	76.8	3.5	1359	4 US-09-833-381-997	Sequence 997, Appli
C 2	74.2	3.4	786431	4 US-09-751-389-3	Sequence 3, Appli
C 3	73.8	3.4	9377	4 US-09-801-874-3	Sequence 3, Appli
C 4	71	3.2	99916	4 US-09-816-095-3	Sequence 3, Appli
C 5	69.2	3.2	694	1 US-08-358-171-25	Sequence 25, Appli
C 6	69.2	3.2	694	3 US-09-090-947-25	Sequence 25, Appli
C 7	69.2	3.2	1201	1 US-08-358-171-9	Sequence 9, Appli
C 8	69.2	3.2	1201	3 US-09-090-947-9	Sequence 9, Appli
C 9	69.2	3.2	1426	1 US-08-358-171-3	Sequence 3, Appli
C 10	69.2	3.2	1426	3 US-09-090-947-3	Sequence 3, Appli
C 11	65.6	3.0	4078	4 US-09-495-050A-204	Sequence 204, Appl
C 12	64.6	2.9	392000	4 US-10-027-983-11	Sequence 11, Appl
C 13	64	2.9	13187	4 US-09-422-936-61	Sequence 61, Appli
C 14	63.2	2.9	81001	4 US-09-750-580-1	Sequence 1, Appli
C 15	61	2.8	84495	4 US-09-797-906-3	Sequence 3, Appli
C 16	60.4	2.8	72928	3 US-09-009-913-1	Sequence 1, Appli
C 17	59.8	2.7	38564	4 US-09-734-673-3	Sequence 3, Appli
C 18	59.4	2.7	44848	4 US-09-435-739-42	Sequence 42, Appli
C 19	58	2.6	168575	4 US-09-426-290-1	Sequence 1, Appli
C 20	56.8	2.6	43795	3 US-08-742-185-101	Sequence 101, Appl
C 21	55.6	2.5	1849	4 US-09-023-655-1232	Sequence 1232, Ap
C 22	55.6	2.5	246240	2 US-08-724-394A-20	Sequence 20, Appl
C 23	55.6	2.5	246240	2 US-08-724-394A-21	Sequence 21, Appl
C 24	55.6	2.5	246240	2 US-08-724-394A-22	Sequence 22, Appl
C 25	54.6	2.5	2012	1 US-07-866-560-3	Sequence 3, Appli
C 26	54.6	2.5	2012	1 US-08-077-673-3	Sequence 3, Appli
C 27	54.6	2.5	2012	1 US-08-478-992-3	Sequence 3, Appli

28	54.6	2.5	2012	3 US-09-105-298-3	Sequence 3, Appli
29	54.6	2.5	2012	3 US-08-706-281A-7	Sequence 7, Appli
30	54.6	2.5	2012	3 US-09-097-231-7	Sequence 7, Appli
31	54.6	2.5	2012	4 US-09-542-122-3	Sequence 7, Appli
32	54.6	2.5	2012	4 US-09-353-099-7	Sequence 7, Appli
33	53.8	2.5	319608	4 US-09-539-333D-1	Sequence 1, Appli
34	53.8	2.5	319608	4 US-09-679-409-1	Sequence 1, Appli
35	52.8	2.4	116592	4 US-09-818-512-3	Sequence 3, Appli
36	51.4	2.3	1001	4 US-09-671-317-355	Sequence 355, App
37	51.4	2.3	1001	4 US-09-671-317-357	Sequence 357, App
38	51	2.3	90541	4 US-09-759-359A-3	Sequence 3, Appli
39	49.2	2.2	695	4 US-09-200-934-6	Sequence 6, Appli
40	49	2.2	87350	3 US-08-781-891-79	Sequence 79, Appl
41	49	2.2	87350	4 US-09-618-166-79	Sequence 79, Appl
42	49	2.2	87543	4 US-09-791-211-3	Sequence 3, Appli
43	48.4	2.2	55827	4 US-09-813-133A-3	Sequence 3, Appli
44	48	2.2	4011	1 US-08-121-057-3	Sequence 3, Appli
45	48	2.2	4011	2 US-08-509-187D-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-833-381-997/c
; Sequence 997, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 997
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-997

Query Match Similarity	3.5%	Score	76.8	DB	4	Length	1359
Best Local Similarity	72.4%	Pred. No.	1.2e-11				
Matches	113	Conservative	0	Mismatches	42	Indels	1
Gaps	1						
QY	1460	ACTTATTGTTGTTTATATACATTTATGAGATACAGTGTAAATTCGTTACATGGATATAT	1519				
DB	1315	ATTTATTATTGTAATACATTTAAAGGGTACAGCATAATTTTGTTCATGGATATAT	1256				
QY	1520	GCCATATTGGTGAAGTCAGAGATTTCAGTGTGCACATCACCGGAAATGTTTAACTGAC	1579				
DB	1255	CACAAAGTGGAGAGCTGTGAGCTTTAGTTTATCCATCACCTAAATATTGTACATTATAC	1196				
QY	1580	CCATTAAAGTAAATTTCTCATCCCATTTCCCTCA	1615				
DB	1195	CCAGT-AGGTAATTTCTCATCCCATTTCCCTCA	1161				

RESULT 2

US-09-751-389-3/c
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match      3.4%; Score 74.2; DB 4; Length 786431;
Best Local Similarity 65.0%; Pred. No. 2.1e-09;
Matches 141; Conservative 0; Mismatches 73; Indels 3; Gaps 2;

QY 1446 TTTGTAAGAAAAAAGCTTTTGTGTTATATACATTTATGAGACAAAGTGTAAATTCG 1505
DB 170208 TTTTAAATGATAAATCTTTTGTACTTGTATTAATGATAGATACAGTATATTTTA 170149

QY 1506 TTACATGGATATATGACATATTGGTGAAGTCAGAGATTTTCAGTGTGCACATCACCGGAAA 1565
DB 170148 TTACACTAATATATTCATATTGGTGAAGTCAGGCGCTTT-AGTGTACTCATCACCGGAAT 170090

QY 1566 AATGTTAACTGTACCATTAAGTAATTTCTCATCCCCCATTTCCCTCACCACATGCTC 1625
DB 170089 AATGTACATTGCACC--ATTAAGTAATTTCTCATCATCATCCCCCTTTTACCCCTTCAC 170032

QY 1626 CAGCCCTTCCAGGAACTGTTCATTTTCTCTGACTA 1662
DB 170031 CTTTCCAAGTCTCCATGTCTATCATTCACAGTGTA 169995

RESULT 3
US-09-801-874-3
; Sequence 3, Application US/09801874
; Patent No. 6582935
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN AMINOTRANSFERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: AMINOTRANSFERASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO00615
; CURRENT APPLICATION NUMBER: US/09/801,874
; CURRENT FILING DATE: 8001-03-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9377
; TYPE: DNA
; ORGANISM: Human
US-09-801-874-3

Query Match      3.4%; Score 73.8; DB 4; Length 9377;
Best Local Similarity 70.2%; Pred. No. 2.4e-10;
Matches 113; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

QY 1462 TTATTTGTTGTTATATACATTTATGAGATACAAAGTGTAAATTCGTTACATGATATATGC 1521
DB 1570 TTTTTTTAAATTTGTATAAATGTGTGAAGTACAAAGTGTAAATTTGTTAGATGCATACATCG 1629

QY 1522 CATATGCGTGAAGTCAGAGATTTTCAGTGTGCATCATCCCCGAAATGTTAACTGTACCC 1581
DB 1630 CATAGTGGTGAAGTCAGGCGCTTTTAGGGTATCCATACCCCAACAATGTACATTGTATCT 1689

QY 1582 ATTAAGTAATTTCTCATCCCCCATTTCCCTCACCACATG 1622
DB 1690 GTT-AGTAATCTCCCATATCTCTGCGCTGTATCTAAGG 1729

RESULT 4
US-09-816-095-3
; Sequence 3, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
; APPLICANT: GAN, Weinliu
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01147
; CURRENT APPLICATION NUMBER: US/09/816,095
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 99916
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(99916)
; OTHER INFORMATION: n = A,T,C or G
US-09-816-095-3

Query Match      3.2%; Score 71; DB 4; Length 99916;
Best Local Similarity 55.5%; Pred. No. 5.5e-09;
Matches 137; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 1362 AGGGAATATAAGTCTCTAATGACTCACAATGTAAGTCTGCTGGGTAAGAAAGTTCTAATGT 1421
DB 34348 AGTGTTTCATAGTTTATTTAGTGTATATGTTTGTGCTCTCTGTTAAGTTATTTATGAT 34407

QY 1422 TGGAACTCCTTTTCAGCTATCGAAATTTTGTAAAGAAAAAACTTTATTTGTTTATATACAT 1481
DB 34408 TTTTATGCTGGTAGGTTTATTTTAAATTTTGGGGAAGTGTTTTATTATAACAAT 34467

QY 1482 TTATGAGATACAAGTGTAAATTTTCGTTACATGATATATGCCATATTTGTTAGTGTGAGT 1541
DB 34468 GTATGGGTTACAAAGTGCAATTTTGTATCATGTATATAGTTGTATAGTGTGATGTCAGGAC 34527

QY 1542 TTTTCAGTGTGCATCATCCCCGAAATGTTAACTGTAACTTACCATTAAAGTAATTTCTCATCC 1601
DB 34528 TTTTATGTTATCCATCATCACTCAATAAATCTTCCATTGTCATCCATTAGTACACCCACTCCC 34587

QY 1602 CCCATTT 1608
DB 34588 ACCTGTT 34594

RESULT 5
US-08-358-171-25
; Sequence 25, Application US/08358171
; Patent No. 5763578
; GENERAL INFORMATION:
; APPLICANT: FONG, Henry K.W.
; TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THEREO
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,171
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
```

REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
OTHER INFORMATION: /note= Exon 7 corresponds to
OTHER INFORMATION: nucleotides 39...670
US-08-358-171-25

Query Match 3.2%; Score 69.2; DB 1; Length 694;
Best Local Similarity 72.5%; Pred. No. 1.2e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1476 ATACATTTATGAGATACAAGTGAATTTTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
DB 423 ATAAATGTAAGGGGTACAGTGCAGTTTGTTCATGATAGATTCCTAGTGGTGAAGT 482

QY 1536 CAGAGATTTTCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCCATTAAAGTAAATTC 1595
DB 483 CTGGGCTTTTAGTGAACATCACCTTAATATATACGTTGTACCCATT-AAGTTATTTTC 541

QY 1596 TCATCCCCCAATTCCTCCACC 1617
DB 542 TCATCCCTCACCCCTCCACC 563

RESULT 6
US-09-050-947-25
Sequence 25, Application US/09090947
Patent No. 6008338
GENERAL INFORMATION:
APPLICANT: FONG, Henry K.W.
TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/090,947
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 08/358,171
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 39...797
US-08-358-171-9

LENGTH: 694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
OTHER INFORMATION: /note= Exon 7 corresponds to
OTHER INFORMATION: nucleotides 39...670
US-09-090-947-25

Query Match 3.2%; Score 69.2; DB 3; Length 694;
Best Local Similarity 72.5%; Pred. No. 1.2e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1476 ATACATTTATGAGATACAAGTGAATTTTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
DB 423 ATAAATGTAAGGGGTACAGTGCAGTTTGTTCATGATAGATTCCTAGTGGTGAAGT 482

QY 1536 CAGAGATTTTCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCCATTAAAGTAAATTC 1595
DB 483 CTGGGCTTTTAGTGAACATCACCTTAATATATACGTTGTACCCATT-AAGTTATTTTC 541

QY 1596 TCATCCCCCAATTCCTCCACC 1617
DB 542 TCATCCCTCACCCCTCCACC 563

RESULT 7
US-08-358-171-9
Sequence 9, Application US/08358171
Patent No. 5763578
GENERAL INFORMATION:
APPLICANT: FONG, Henry K.W.
TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,171
FILING DATE: 16-DEC-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 39...797
US-08-358-171-9

Query Match 3.2%; Score 69.2; DB 1; Length 1201;

Best Local Similarity 72.5%; Pred. No. 1.6e-09; Mismatches 38; Indels 1; Gaps 1;
Matches 103; Conservative 0;
QY 1476 ATACATTTATGAGATACAAAGTGTAAATTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
Db 1053 ATAAATGTAAGGGGTACAGTGCAGTTTGTACATGATAGATTGCCCTAGTGGTGAAGT 1112
QY 1536 CAGAGATTTAGTGTGACATACACCCGAAAAATGTTAACTGTACCATTAAGTAATTTTC 1595
Db 1113 CTGGGCTTTTAGTGTAAACCATCACCCCTAATAATATACGTTGTACCCATT-AAGTTATTTC 1171
QY 1596 TCATCCCTCACCCTCCACC 1617
Db 1172 TCATCCCTCACCCTCCACC 1193

RESULT 8
US-09-090-947-9
; Sequence 9, Application US/09090947
; Patent No. 6008338
; GENERAL INFORMATION:
; APPLICANT: FONG, Henry K.W.
; TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/090,947
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/358,171
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: FONG=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 39..797
US-09-090-947-9
Query Match 3.2%; Score 69.2; DB 3; Length 1201;
Best Local Similarity 72.5%; Pred. No. 1.6e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1476 ATACATTTATGAGATACAAAGTGTAAATTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
Db 1053 ATAAATGTAAGGGGTACAGTGCAGTTTGTACATGATAGATTGCCCTAGTGGTGAAGT 1112
QY 1536 CAGAGATTTAGTGTGACATACACCCGAAAAATGTTAACTGTACCCATTAAAGTAATTTTC 1595

Db 1113 CTGGGCTTTTAGTGTAAACCATCACCCCTAATAATATACGTTGTACCCATT-AAGTTATTTC 1171
QY 1596 TCATCCCTCACCCTCCACC 1617
Db 1172 TCATCCCTCACCCTCCACC 1193

RESULT 9
US-08-358-171-3
; Sequence 3, Application US/08358171
; Patent No. 5763578
; GENERAL INFORMATION:
; APPLICANT: FONG, Henry K.W.
; TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,171
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: FONG=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 39..911
US-08-358-171-3
Query Match 3.2%; Score 69.2; DB 1; Length 1426;
Best Local Similarity 72.5%; Pred. No. 1.7e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1476 ATACATTTATGAGATACAAAGTGTAAATTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
Db 1167 ATAAATGTAAGGGGTACAGTGCAGTTTGTACATGATAGATTGCCCTAGTGGTGAAGT 1226
QY 1536 CAGAGATTTAGTGTGACATACACCCGAAAAATGTTAACTGTACCCATTAAAGTAATTTTC 1595
Db 1227 CTGGGCTTTTAGTGTAAACCATCACCCCTAATAATATACGTTGTACCCATT-AAGTTATTTC 1285
QY 1596 TCATCCCTCACCCTCCACC 1617
Db 1286 TCATCCCTCACCCTCCACC 1307

RESULT 10
US-09-090-947-3
; Sequence 3, Application US/09090947

Patent No. 6008338
GENERAL INFORMATION:
APPLICANT: FONG, Henry K.W.
TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/090,947
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/358,171
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 39...911
US-09-090-947-3

Query Match 3.2%; Score 69.2; DB 3; Length 1426;
Best Local Similarity 72.5%; Pred. No. 1.7e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
Qy 1476 ATACATTATGAGATACACAGTGTAAATTCGTACATGATATATGCGCATATATGCGTGAAGT 1535
Db 1167 ATAATGTAAAGGGGTACAGTGCAGTTTGTGTACATGATAGATTGCTGTGGAAGT 1226
Qy 1536 CAGAGATTTCAGTGTGCATACACCCGAAATTTAACTGTACCCATTAAGTAATTC 1595
Db 1227 CTGGCTTTTAGTGAACATCACCCCTAATATATACCTGTACCAATT-AGTTATTC 1285
Qy 1596 TCATCCCCCATTCCTCCCTCACC 1617
Db 1286 TCATCCCTCACCCCTCCACC 1307

RESULT 11
US-09-495-050A-204
Sequence 204, Application US/09495050A
Patent No. 6492505
GENERAL INFORMATION:
APPLICANT: Roopa, Reddy
APPLICANT: Guegler, Karl, J.
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED
FILE REFERENCE: PA-0013 US
CURRENT APPLICATION NUMBER: US/09/495,050A

CURRENT FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/118,318
PRIOR FILING DATE: February 1, 1999
NUMBER OF SEQ ID NOS: 305
SOFTWARE: PERL Program
SEQ ID NO 204
LENGTH: 478
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6492505 2375491CT1
US-09-495-050A-204

Query Match 3.0%; Score 65.6; DB 4; Length 478;
Best Local Similarity 51.9%; Pred. No. 9.9e-09;
Matches 104; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 1450 TAAGAAAAAACTTAATTTGTTTATATATACATTTATGAGATACAAAGTGTAAATTCCTTAC 1509
Db 263 TAAATTTTATTTTACTTTTTCAGTATAAATTTAAGGGGAACAAGTGCAGTTTCTTAC 322
Qy 1510 ATCGATATATGCCATATATGTTGTAAGTCAGAGATTTGAGTGTGCATCACCACCCGAAAATG 1569
Db 323 ATGGATATATTTGCTAGTGTGAAGCTGGGCCCTTTAATGGGACCATCGCCCGGATAGTG 382
Qy 1570 TTAACCTGTACCCATTAAAGTAATTTCTCATCCCCCATTTCCCTCACC 1617
Db 383 TGCATTGTACCCCATATGAATTTACTCAACCCCTACTCACCCTACCACC 430

RESULT 12
US-10-027-983-11
Sequence 11, Application US/10027983
Patent No. 6617162
GENERAL INFORMATION:
APPLICANT: Kenneth W. Dobie
APPLICANT: Mark P. Roach
TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
FILE REFERENCE: RFS-0340
CURRENT APPLICATION NUMBER: US/10/027,983
CURRENT FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 11
LENGTH: 392000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 137740
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 137742
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (138122)...(138221)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 145507
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 151967
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (151967)...(1542066)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 154217
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (164037)...(164136)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc_feature

LOCATION: (8312)...(8408)
NAME/KEY: exon
LOCATION: (8409)...(8516)
NAME/KEY: intron
LOCATION: (8517)...(8928)
NAME/KEY: exon
LOCATION: (8929)...(9022)
NAME/KEY: intron
LOCATION: (9023)...(9716)
NAME/KEY: exon
LOCATION: (9717)...(9844)
NAME/KEY: intron
LOCATION: (9845)...(10424)
NAME/KEY: exon
LOCATION: (10425)...(10553)
NAME/KEY: intron
LOCATION: (10554)...(10775)
NAME/KEY: exon
LOCATION: (10776)...(10919)
NAME/KEY: intron
LOCATION: (10920)...(11700)
NAME/KEY: exon
LOCATION: (11701)...(13187)
US-09-422-936-61

Query Match 2.98; Score 64; DB 4; Length 13187;
Best local Similarity 68.8%; Pred. No. 1.7e-07;
Matches 88; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1458 AAATTATTTGTTTATATACATTTATGAGATACAAAGTGTAAATTCGTTACATGGATAT 1517
DB 4960 AAAAATTTCTATTTTATGTAATTTAGAGGTAGAGTGCAGTTGTTTACATGAATAT 5019

QY 1518 ATGCATATTGGTGAAGTCAGAGATTTTCAGTGTGCACATCACCCGAAAATGTTAACTGT 1577
DB 5020 ATTGCATAGTGTGAAGTCCGGGGTTTGTAGTGCCTGTCCCGAACAGTGCACCTTGT 5079

QY 1578 ACCCAATTA 1585
DB 5080 ACCTAATA 5087

RESULT 14

US-09-750-580-1
Sequence 1, Application US/09750580
Patent No. 6455280
GENERAL INFORMATION:
APPLICANT: Yen, Frances
APPLICANT: Denison, Blake
APPLICANT: Bour, Barbara
APPLICANT: Bhaïm, Bernard
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Duclert, Aymeric
APPLICANT: Bougueleret, Lydie
APPLICANT: Ebbets-Reed, Dana
APPLICANT: Salter-Cid, Luisa
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
FILE REFERENCE: 89.US2.CIP
CURRENT APPLICATION NUMBER: US/09/750,580
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 09/599,362
PRIOR FILING DATE: 2000-06-21
PRIOR APPLICATION NUMBER: PCT/TB00/0101
PRIOR FILING DATE: 2000-06-21
PRIOR APPLICATION NUMBER: PCT/TB99/02058
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: US 49/469/099
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: US 60/113,686
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: US 60/141,032
PRIOR FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 6

SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 81001
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 10946..12946
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 12947..12958
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 13470..13526
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 13641..13752
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 14271..15968
OTHER INFORMATION: exon 4
NAME/KEY: misc feature
LOCATION: 15969..17969
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1239
OTHER INFORMATION: 20-828-311 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12347
OTHER INFORMATION: 17-42-319 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 15241
OTHER INFORMATION: 17-41-250 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 42218
OTHER INFORMATION: 20-841-149 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 45442
OTHER INFORMATION: 20-842-115 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 77058
OTHER INFORMATION: 20-853-415 : polymorphic base C or T
NAME/KEY: primer bind
LOCATION: 929..949
OTHER INFORMATION: 20-828.pu
NAME/KEY: primer bind
LOCATION: 1357..1377
OTHER INFORMATION: 20-828.rp complement
NAME/KEY: primer bind
LOCATION: 12029..12050
OTHER INFORMATION: 17-42.pu
NAME/KEY: primer bind
LOCATION: 12581..12603
OTHER INFORMATION: 17-42.rp complement
NAME/KEY: primer bind
LOCATION: 14992..15012
OTHER INFORMATION: 17-41.pu
NAME/KEY: primer bind
LOCATION: 15460..15482
OTHER INFORMATION: 17-41.rp complement
NAME/KEY: primer bind
LOCATION: 42070..42090
OTHER INFORMATION: 20-841.pu
NAME/KEY: primer bind
LOCATION: 42572..42591
OTHER INFORMATION: 20-841.rp complement
NAME/KEY: primer bind
LOCATION: 45328..45347
OTHER INFORMATION: 20-842.pu
NAME/KEY: primer bind
LOCATION: 45863..45883
OTHER INFORMATION: 20-842.rp complement
NAME/KEY: primer bind

Mon Mar 29 09:55:00 2004

```

LOCATION: 76644..76664
OTHER INFORMATION: 20-853.pu
NAME/KEY: primer bind
LOCATION: 77166..77185
OTHER INFORMATION: 20-853.rp complement
NAME/KEY: primer bind
LOCATION: 1220..1238
OTHER INFORMATION: 20-828-311.mis
NAME/KEY: primer bind
LOCATION: 1240..1258
OTHER INFORMATION: 20-828-311.mis complement
NAME/KEY: primer bind
LOCATION: 12328..12346
OTHER INFORMATION: 17-42-319.mis
NAME/KEY: primer bind
LOCATION: 12348..12366
OTHER INFORMATION: 17-42-319.mis complement
NAME/KEY: primer bind
LOCATION: 15222..15240
OTHER INFORMATION: 17-41-250.mis
NAME/KEY: primer bind
LOCATION: 15242..15260
OTHER INFORMATION: 17-41-250.mis complement
NAME/KEY: primer bind
LOCATION: 42199..42217
OTHER INFORMATION: 20-841-149.mis
NAME/KEY: primer bind
LOCATION: 42219..42237
OTHER INFORMATION: 20-841-149.mis complement
NAME/KEY: primer bind
LOCATION: 45423..45441
OTHER INFORMATION: 20-842-115.mis
NAME/KEY: primer bind
LOCATION: 45443..45461
OTHER INFORMATION: 20-842-115.mis complement
NAME/KEY: primer bind
LOCATION: 77039..77057
OTHER INFORMATION: 20-853-415.mis
NAME/KEY: primer bind
LOCATION: 77059..77077
OTHER INFORMATION: 20-853-415.mis complement
NAME/KEY: misc binding
LOCATION: 1227..1251
OTHER INFORMATION: 20-828-311.probe
NAME/KEY: misc binding
LOCATION: 12335..12359
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc binding
LOCATION: 15229..15253
OTHER INFORMATION: 17-41-250.probe
NAME/KEY: misc binding
LOCATION: 42206..42230
OTHER INFORMATION: 20-841-149.probe
NAME/KEY: misc binding
LOCATION: 45430..45454
OTHER INFORMATION: 20-842-115.probe
NAME/KEY: misc binding
LOCATION: 77046..77070
OTHER INFORMATION: 20-853-415.probe
US-09-750-580-1

```

```

1565  AAATGTTAACTAGCCATTAAAGTAATTTCTCATCCCCCATTTCCCTCCACCACTGCT 1624
Db      67610  TAATGTACATAAAGGCCCA-TGAAGTAATTTCTCATCATTCACCCCTCCACCCCTCTCA 67668

QY      1625  CCAGCCCTTGCC 1636
Db      67669  CCCTTCCAGTC 67680

RESULT 15
US-09-797-906-3
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: Zianghe YAN, Karen A. KETCHUM, Valentina DiFRANCESCO, Ellen M. BEN
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001151CIP
; CURRENT APPLICATION NUMBER: US/09/797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 84495
; TYPE: DNA Human
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(84495)
; OTHER INFORMATION: n = A,T,C or G
; US-09-797-906-3

```

Search completed: March 28, 2004, 05:41:05
Job time : 193 secs

Result	Query	Score	Length	DB	ID	Description
No.	Match					

1	240	11.0	2709	15	US-10-133-937-84	Sequence 84, Appl
2	240	11.0	2709	15	US-10-373-801-17	Sequence 17, Appl
3	240	11.0	2709	15	US-10-159-563-84	Sequence 84, Appl
4	238.4	10.9	1254	14	US-10-084-817-69	Sequence 69, Appl
5	238.4	10.9	2716	14	US-10-240-965-259	Sequence 259, App
6	228.8	10.4	2736	15	US-10-264-494-498	Sequence 498, App
C 7	165.2	7.5	15732	14	US-10-239-676-96	Sequence 96, Appl
C 8	165.2	7.5	15732	14	US-10-240-453-108	Sequence 108, App
9	157.4	7.2	15732	14	US-10-239-676-95	Sequence 95, Appl
10	157.4	7.2	15732	14	US-10-240-453-107	Sequence 107, App
11	157	7.2	2070	15	US-10-108-260A-450	Sequence 450, App
12	157	7.2	2905	13	US-10-002-600-11	Sequence 11, Appl
13	157	7.2	2905	14	US-10-084-817-70	Sequence 70, Appl
14	157	7.2	2906	10	US-09-971-392-88	Sequence 88, Appl
15	106.4	4.9	1892	12	US-10-433-544-13	Sequence 13, Appl

001169261733YH3 06047 6X

[illegible]

QY	562	CCAGATCCCTCTTTGGCTTTTCTTGACTTACTCTCCTTTTGTGATAAAGTATTTCATATGGTT	621
Db	11813	TCATAATCCTCTAAATTTTCTAAATTTACCCCTACTTTTAATAAATAATTTAAACA--ATT	11757
QY	622	TCAAATCCAGAAATTCAGCTATTTCACCTCATGGCCCTAGTTTGGATTAAAACAATCAGCTT	681
Db	11756	CCAAATCCAAACTTTA--ACTCTCTCTATACTACTAAATTTAAATTTAAAACAATCAACTT	11700
QY	682	TCCTTTGCTCTGTAGATGGCTTTGGACAGT-----ACATATTAAACAGTCTATGGCCCTT	736
Db	11699	TATTTCTTCTATAAATAACCTTTAAATCAATTTTCCCTATTAACTATTTTAAAATCCCTT	11640
QY	737	TCACCTCATGACCAATCCTGCTTGCAGATCTGCTGGGTAAACACCAATCATCTGC	796
Db	11639	TCCAATCATAAACCAATTTTACTTACAAATCAACTAAATTAATAACCAACAACATATCTAC	11580
QY	797	TAGTTGATGGCAATTAATTTGACATCTCTGGGGTTACAGCAGCTTAAGGGCAACCATGTGC	856
Db	11579	TTACTAATACCAATTCATTTTAACAATTTCAAAATTTATAATACCTTTTAATACACCCACTAT	11520
QY	857	CTTAAACAGAAAAATACACTGCTGAGCAGAGGATTAACAGACCTCCACACTACCGGA	916
Db	11519	CTTCACAAA--AAAAAAAATTAACAACAAAAATTTATCAAAACCCCTTATCATTTCCAAA	11461
QY	917	AGTTGAGTGCATTTTCCAGGATTTTCAAGTCTTCGATGCTGCATGTTTGAGAGTCCACTA	976
Db	11460	ATTTCAATAC-TATTCAAAAATTTCAAAATTTCAATCAATATAAACCBA-----	11410
QY	977	TCCACGTTTTGTTTATCCAGGAATAATGCAGAAAACTGAGCANTCGTGAATCTAATCAT	1036
Db	11409	-----CTATCCATAATAATACAAAAAACAATAAATAATCATAAATTTTAATAATC	11361
QY	1037	GAGGTAAAGTAACCTAGTCAAG--TCAAGAGGACAGCTAGTGTAACTTCCTTCATAT	1093
Db	11360	AAATTAACAACACTCAATATTATATCTCAAAAAAACAATAATAAANAATCTTTCAAA	11301
QY	1094	AAATTTAAGGAAGATTTACATATTCATAATGCATTTCT--CATCTGGAAAGCCAGGCT	1151
Db	11300	-TATTTAAATAAANAATTTACATATTCATAATAACTTTTACCTCTATCAACTCCAAATA	11242
QY	1152	TGTCAGTTATTAATAATTTGGATGAATCTCAGGTAGTATTTAAAAACAATAGCATTTCA	1211
Db	11241	TATCAATATTATACTAAATAAATA-TATAAATACATTTTAAACAAA-AACATTTCA	11184
QY	1212	TAGAAGACATTTCCCAACACAGAGATGACGTAAGTATTGGTAGCAATAGATGCTCTA	1271
Db	11183	CAATAAACCCACCCCTATCCMAAAAAATAAATCAAAATTTCTAATACAATAAACAATCTCTA	11124
QY	1272	CCAAGATCTCCTCTGGCACTGCCACTCTGTGCACAGGAAGAAGGGGGAAGAGCCAT	1331
Db	11123	CCAAAATCTCTACTCTCTACCACCTPATAAACAAAAAAAATAAATACTTTTAATATTT	11064
QY	1332	GTTCCTCTCTTATATTG-----AGCATCATTTAAGGGAATATAAGTCTCTAATGACTCA	1387
Db	11063	TACCTTTCCCAATTTATTTAAATCCTTATTCAAACGACTACCGATTTCTAATAACCTA	11004
QY	1388	CATGTAAGTCTGCTGGGTAAAGAAAGTTCTAAATGTGGAACTCCTTTTACGTATCGAATTT	1447
Db	11003	TATATAAATTAACCTTTTATAAAAAAATTTCTAAATTTTAAACTCTCTTTCACAAATATTTAT	10944
QY	1448	TGTAAGAAAAA 1459	
Db	10943	AAATCAAAAAA 10932	

RESULTS

US-10-240-453-108/c
; Sequence 108, Application US10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian

Db 11360 AAAATAAAACAACTAATTAATTAATCTCAAAAACAACTAATTAATTAATTAATCTTTCAAT 11301
 QY 1094 AAATTTAAGGAAGATTTACATATTCATATGACTTTCT--GATCTGGAAGCCAGGCT 1151
 Db 11300 -TATTTAATAAATAATTTACATATTCATATTAATCTTTACCTCTATCAAACTCCAAATA 11242
 QY 1152 TGTCAAGTATTAATTAATTTGATGAATACATCTCAGGTAGTATTTAAACAATAGCATTTCA 1211
 Db 11241 TATCAATTAATTAATACTAATAATAA-TATAAATACTATTTAAACAAA-AAACATTTCA 11184
 QY 1212 TAGAAGACATTTCCCAACACAGAGATGTCAGTAAAGTAATTTGGTAGCAATAGATGCTCCTA 1271
 Db 11183 CAATAAACCCACCTCTATCCAAAATAAATAATCAATTTCTAATAAACAATAACATCTCTA 11124
 QY 1272 CCAAGATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1331
 Db 11123 CCAAAATCT 11064
 QY 1332 GTTGCTTCTCTTATTTTG---AGCATCATTTAAGGGAATATAAGTCTCTAATGACTCA 1387
 Db 11063 TACCTTTCCCAATTTATATTAATAATCTCTTATCAACGACTACCGATTTCTAATACTTA 11004
 QY 1388 CATCTAAGTCTGCTGGTGAAGATTTCTAATGTTGGAATCTCTTTTCCAGCTATGCAATTT 1447
 Db 11003 TATAAATAATTAATTTAATAAATAATTTCTAATAATTTAATAATTTCTCTTCAACAATAATTTAT 10944
 QY 1448 TGTAAGAAAAAA 1459
 Db 10943 AAATCAAAAAAA 10932

RESULT 9

US-10-239-676-95
 ; Sequence 95, Application US/10239676
 ; Publication No. US20030082609A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
 ; FILE REFERENCE: 5013.1003
 ; CURRENT APPLICATION NUMBER: US/10/239,676
 ; CURRENT FILING DATE: 2002-09-24
 ; PRIOR APPLICATION NUMBER: PCI/EP01/03968
 ; DE 10019058.8
 ; DE 10019173.8
 ; DE 10032529.7
 ; DE 10043826.1
 ; PRIOR FILING DATE: 2001-04-06
 ; 2000-04-06
 ; 2000-04-07
 ; 2000-06-30
 ; 2000-09-01
 ; NUMBER OF SEQ ID NOS: 228
 ; SEQ ID NO 95
 ; LENGTH: 15732
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 ; US-10-239-676-95

Query Match 7.2%; Score 157.4; DB 14; Length 15732;
 Best Local Similarity 56.9%; Pred. No. 4.2e-29;
 Matches 526; Conservative 0; Mismatches 351; Indels 48; Gaps 11;
 QY 528 ATCAGAACACCCAGAGATGTTATGGCCATATTCAGATCTCTTTGCTTTTCTTCAC 587
 Db 3886 ATAATGAATATTTATGTTAGTGGTGTATTTAGATTTTGTGTTTGTGTTTGTGAT 3945
 QY 588 TTATCTCTTTTGTGATAAGTATTATCATATGTTTCAATCCAGAAATTCAGCTATTCAC 647
 Db 3946 TTA---TTTGTGTTTAGTGGGTTTAGATAGTTTAAATTTAAATTTTAAATTTT 3999

QY 648 TCATGSCCCTAGTTGGATTAAAAAACAATCAGCTTTCTTTTGTCTGTAGATGCCCTTGA 707
 Db 4000 TTATGTTGTTGGTTGGATTAGATTAATTTAGTTTATTTTGTAGATGTTTGA 4059
 QY 708 CAGTACATATTAACAGTCTPAT-----GGGCCCTTTCCACTCAGACCACTCCTGCTGC 762
 Db 4060 TTATTTTATTTTATTTAGTTGTTTGGGATTTTATTTTATTTAGTATGATTTATTTTGT 4119
 QY 763 AGATCTGCTGCTGGGTTAACACCATCATCTGCTAGTTGATGSCCATTTTATTTGACATCC 822
 Db 4120 AGATTAGTTAGTTGGTAAATTTAGTATTTGTTGTTAAATTTGATATTTATTTTGTAAAT 4179
 QY 823 TGGGGTTTACAGCAGCTGCTAAAGGGCACCAGTGTCTTAAACAGAGAAAAATACACTGCTGA 882
 Db 4180 TAGAGTTATGATATTTTGTAGTGTATTTATTTGTTTATAG-GAAAAAATAATTTAGT 4238
 QY 883 GCAGAGATTTAAACAGACTCTCTCACACTACCGGAAGTTTCACTGCAATTTCCAGGATTTTC 942
 Db 4239 ATAAAGGGTTTATTTAGGTTTATTTATTTTATTTAGAGTTTGTAGTGTGTTTGTAGAT 4298
 QY 943 AAGTCTTCGATGCTGCTCACTGAGAGTCCACTATCCAGTTTGTGTTTATCCAGGAATAAT 1002
 Db 4299 AGTTTATGATGATGAA-----TTAATTTATTAATAAT 4338
 QY 1003 GCAGAAAACTGAGCATTTCTGAAATCTAATCAATGAGTAAAGTAACTAGTCACTGCTCA 1062
 Db 4339 GTAAAAAGTATGATGATTAATTAATTAATTAAGGTAAGGTAGTTTATTTATTTATAT 4398
 QY 1063 AGAGCAGCTAGTGTTA--AACTCCTTCATATAAATTTAAGGAAGATTTACATATTTCA 1120
 Db 4399 TTAAGAGAGTAAATTAGTATGAAATTTTAAATTTTAAATTTTGAATTAAGATTTATTTA 4458
 QY 1121 TAATGACTTTCTGATCTGG--AAAGCCAGGCTTGTGAGTTTATTAATAATTTGATCAATA 1178
 Db 4459 TAATGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 4518
 QY 1179 CTTCAAGTATGATTTTAAACAAATAGGATTTTATAGAGACATTTCCCAACACAGAGAT 1238
 Db 4519 -GTAGGTTATTTTAAATAAA-AAATAATTTTATAGTAGATTTATTTTATTTAAGAGAT 4576
 QY 1239 GCAGTAAAGTATTTGTTAGCAATAGATCTCTCTTACCAAGATCTCTCTGCTGCTGCT 1298
 Db 4577 GAAGTTAAAGTTTGTGTTAGTAAATGATATTTTATTAAGAT---TTTTGTTTGTGTTAT 4633
 QY 1299 CTGTGCACAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1351
 Db 4634 TTGTGAATAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4693
 QY 1352 GCATCATTTAAGGGAATATAAGTCTCTAATGACTCAATGATGCTGCTGCTGCTGCTGCT 1411
 Db 4694 TTTTATTTAAGCGTTATCGGTTTTTAAATGATTTATATGATGATGATGATGATGATGATG 4753
 QY 1412 GTTCTAATGTTGAACTCCTTTTCTAG 1436
 Db 4754 GTTTTGTAGTTTGAATTTTATTTTAG 4778

RESULT 10

US-10-240-453-107
 ; Sequence 107, Application US/10240453
 ; Publication No. US20030148326A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
 ; TITLE OF INVENTION: Transcription
 ; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
 ; TITLE OF INVENTION: With DNA Transcription
 ; FILE REFERENCE: 5013.1009
 ; CURRENT APPLICATION NUMBER: US/10/240,453
 ; CURRENT FILING DATE: 2002-10-02

PRIOR APPLICATION NUMBER: PCT/EP01/03973
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 350
SEQ ID NO 107
LENGTH: 15732
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-107

Query Match 7.2%; Score 157.4; DB 14; Length 15732;
Best Local Similarity 56.9%; Pred. No. 4.2e-29;
Matches 526; Conservative 0; Mismatches 351; Indels 48; Gaps 11;

QY 528 ATCCAGAACCCAGAGATGTTATGGCCATATCCAGATCTCTTTGCTTTTCTGAC 587
DB ATATGAATATTTATATGTTAGTGGTGTGTTATTTAGATTTTGTGTTTGTGAT 3945
QY 588 TTACTCTTTTGTGATAAGTTATTCATATGTTTCAATCCAGATTCAGTATCTACT 647
DB TTA-----TTTGTGTTTGTAGTGGGTTTAGATATTTTAAATTTAAATTT- -TTAATTTTTT 3999
QY 648 TCATGGCCCTAGTTGGATTAACAACATCAGCTTCTTTTCTGCTGTAGATGGCCCTCGA 707
DB TTATGTTGTTGGGTTTGGATTAAGATAATAGTTTATTTTGTGATGATTTATTTGTTAT 4059
QY 708 CAGTACATATTAAACAGTCTAT-----GGGCCCTTTCCACTCATGACCACCTCTCTGCG 762
DB TTAATTTTTTTTATAGTTGTTTGGGATTTTTTTTAGTTATGATTTATTTTGTGTTAT 4119
QY 763 AGATCTGCTGCTGGTTTAACACCAATCATCTGCTAGTTGATGCCATTTATTTGACATCC 822
DB AGATTAGTTAGTTGCTAAATATATAGTATTTGTTTAAATGATATTTATTTTGTGATAAT 4179
QY 823 TGGGTTACAGCACCTAAAGGSCACCCAGTCTCTTAACAGAGAAAAATACACGCTGA 882
DB TAGAGTTATGATATTTTGTAGTGTATTTATGTTTTTTTATAG-GAAAAAAATTTGTAGT 4238
QY 883 GCAGAGGATTAACAGACTCTCCACACTACCGGAAGTTCAAGTCAATTTCCAGGATTTTC 942
DB ATAAAGGTTTATTAGGTTTTTTTATTATTTTAGAGTTTGTAGTGTGTTTAGAATGTTTA 4298
QY 943 AAGCTTTCAGTCTGTCATGTTGAGAGTCCACTATCCAGGTTTGTGTTATCCAGGAATAT 1002
DB AGTTTATAGTTATGATATGAA-----TTAATTTATTTAATAATAAT 4338
QY 1003 GCAGAAAAATCTGAGCAATTCGTGAATCTAATCATTCAGGTAAAGTAACTAGTCATGGTCA 1062
DB GTAAAAAGTATGAGTAATTAAGAAATTAATAATTAAGTTAGTTAGTTATTTATAT 4398
QY 1063 AGAGGACAGCTAGTGTGA--AACTCCTTCATATAAATTAAGGAAGGATTTACATATCA 1120
DB TTAAGAGAGTAATTAGTATGAAATTTTTTAAATTTATTTGAATAAGGATTTATATATTA 4458
QY 1121 TAACTACTTCTGATCTGG--AAAGCCAGGCTTCTCAGTTATTAATTAATTTGGATGAATA 1178
DB TAAATGATTTTTTGTGTTTGTAGTGTATTTAGTGTATTTAGTTATTAATAATTTGGGTAATA 4518
QY 1179 CTTCAAGGTAGTATTTAAACAAATAGCGATTTTCATAGAAGACATTTCCCAACACGAGAT 1238
DB -TGAGGTATTTATTAATAA--ATAATTTTATAGTAGATTTATTTTATTTAAGAT 4576
QY 1239 GCAGTAAAGTATTTGGTAGCAATAGATGCTCTCCACCAAGATCTCTCTGCCACTGCCACT 1298

DB 4577 GAAGTTAAAGTTTGGTAGTAATAGATATTTTATTAAAGAT---TTTGTGTTTGTGTTAT 4633
QY 1299 CTGTGCACAGAAAGAGGGAGAGAGCCATGTTGCTTCTCTTATATTT-----GA 1351
DB 4634 TTGTGAATAGGAAAGGAGGAGAAATTTTGGTGTGTTTGTGTTTATTTAATTTAATAAGA 4693
QY 1352 GCATCATTTTAAAGGAATATAAGTCTCTAATGACTCACAATGTAAGTCTGCTGGGTAPGAAA 1411
DB 4694 TTTTATTTTAAAGCGTTATCGTTTATTAATGATTTATATGTAAGTTAGTTTATATAGAAA 4753
QY 1412 GTTCTAATGTTGGAACCTCCTTTCCAG 1436
DB 4754 GTTTTGAGTTTGAAATTTTTTTTAG 4778

RESULT 11
US-10-108-260A-450
Sequence 450, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: No. US20040005560A1el full length cdna
FILE REFERENCE: H1-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
CURRENT FILING DATE: 2002-03-27
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 450
LENGTH: 2070
TYPE: DNA
ORGANISM: Homo sapiens
US-10-108-260A-450

Query Match 7.2%; Score 157; DB 15; Length 2070;
Best Local Similarity 100.0%; Pred. No. 1.6e-29;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1813 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACACATCCTCGGTTCCGTTTCTGGG 1872
DB 428 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACACATCCTCGGTTCCGTTTCTGGG 487
QY 1873 AACTTTACTGATTTATCTCCCTCCCTCACAAATAAGCATTTCTCGCATTTCTGAAG 1932
DB 488 AACTTTACTGATTTATCTCCCTCCCTCACAAATAAGCATTTCTCGCATTTCTGAAG 547
QY 1933 ATCTCAGATCTGACTACTCTTGAAAAAATTTCCAG 1969
DB 548 ATCTCAGATCTGACTACTCTTGAAAAAATTTCCAG 584

RESULT 12
US-10-002-600-11
Sequence 11, Application US/10002600
Publication No. US20020137077A1
GENERAL INFORMATION:
APPLICANT: Hopkins, Christopher M.
APPLICANT: Peterson, David P.
APPLICANT: Cocks, Benjamin G.
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
FILE REFERENCE: PA-0042 US
CURRENT APPLICATION NUMBER: US/10/002,600
CURRENT FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: 60/243,521
PRIOR FILING DATE: 2000-10-25
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PERL Program
SEQ ID NO 11
LENGTH: 2905
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature

```
; OTHER INFORMATION: Template ID: 040652.36
US-10-002-600-11
Query Match
Best Local Similarity 7.2%; Score 157; DB 13; Length 2905;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1813 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 1872
DB 107 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 166
QY 1873 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTCTGCAATTCCTGAAG 1932
DB 167 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTCTGCAATTCCTGAAG 226
QY 1933 ATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 1969
DB 227 ATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 263
RESULT 13
US-10-084-817-70
; Sequence 70, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Plon
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: FA-0046 US
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 70
; LENGTH: 2905
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 040652.36
US-10-084-817-70
Query Match
Best Local Similarity 7.2%; Score 157; DB 14; Length 2905;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1813 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 1872
DB 107 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 166
QY 1873 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTCTGCAATTCCTGAAG 1932
DB 167 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTCTGCAATTCCTGAAG 226
QY 1933 ATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 1969
DB 227 ATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 263
```

```
RESULT 14
US-09-971-392-88
; Sequence 88, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:
; APPLICANT: Peterson, David P.
; APPLICANT: Peterson, Cecelia I.
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
; FILE REFERENCE: FA-0029 US
```

```
; CURRENT APPLICATION NUMBER: US/09/971,392
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/237,652
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PERL Program
; SEQ ID NO 88
; LENGTH: 2906
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 040652.29
US-09-971-392-88
Query Match
Best Local Similarity 7.2%; Score 157; DB 10; Length 2906;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1813 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 1872
DB 107 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 166
QY 1873 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTCTGCAATTCCTGAAG 1932
DB 167 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTCTGCAATTCCTGAAG 226
QY 1933 ATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 1969
DB 227 ATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 263
RESULT 15
US-10-433-544-13
; Sequence 13, Application US/10433544
; Publication No. US20040053396A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, Jennifer L.; BAUGHN, Maria R.;
; APPLICANT: YUE, Henry; GIETZEN, Kimberly J.;
; APPLICANT: TANG, Y. Tom; THANGAVELU, Kavitha;
; APPLICANT: LAL, Preeti G.; XU, Yuming;
; APPLICANT: ELLIOTT, Vicki S.; LU, Dyung Aina M.;
; APPLICANT: YAO, Monique G.; LU, Yan;
; APPLICANT: HAFALIA, April J.A.; RAMKUMAR, Javalaxmi
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PF-0858 USN
; CURRENT APPLICATION NUMBER: US/10/433,544
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: PCT/US01/46874
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/251,791
; PRIOR FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 1892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040053396A1 2344051CB1
US-10-433-544-13
Query Match
Best Local Similarity 4.9%; Score 106.4; DB 12; Length 1892;
Matches 135; Conservative 0; Mismatches 26; Indels 3; Gaps 1;
QY 1806 TCTTTGCAGATACCTTCATTTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTT 1865
DB 2 TCTTTGAAATATCTTCATTTTCTTAGCATTTTCAGGAGATTATACATCTCTGATTTTCTAGT 61
QY 1866 TCTTGGGAATCTTACTGATTTATCTCCCGCTCACAAATAAGCATTTCTGCAATTCCTGAAG 1925
```

us-09-931-733-1.rnpb

Mon Mar 29 09:55:01 2004

Db 62 TTCTGAGAGCTTACTGACTGATTTCCTATTCA--AAACAATCCTCATTTCCTACATT 118
Oy 1926 TCTGAAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 1969
Db 119 TCTGAAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 162

Search completed: March 28, 2004, 07:59:35
Job time : 796 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 23:46:19 ; Search time 5729 Seconds
(without alignments)

11420.499 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 2191

Sequence: 1 tatcaaatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_ges_hum:*

18: em_ges_inv:*

19: em_ges_pin:*

20: em_ges_vrt:*

21: em_ges_fun:*

22: em_ges_nam:*

23: em_ges_mus:*

24: em_ges_pro:*

25: em_ges_rod:*

26: em_ges_phg:*

27: em_ges_vrl:*

28: gb_ges1:*

29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244	11.1	300	9 AU100150	AU100150
2	240	11.0	480	9 AL702095	AL702095 DKFZP686A
3	240	11.0	656	13 BU927876	BU927876 AGENCOURT
4	240	11.0	742	14 CF596894	CF596894 AGENCOURT

5	239.4	10.9	497	12	BM819118
6	238.4	10.9	746	14	CD639214
7	237.4	10.8	406	10	AM139342
8	234	10.7	746	14	CD104534
9	233	10.6	608	9	AL711311
10	233	10.6	854	12	BG253409
11	232	10.6	873	14	CD252021
12	232	10.6	1014	12	BM463414
13	231.4	10.6	817	13	BU934276
14	231	10.5	714	12	BG540026
15	223	10.2	579	9	AU279766
16	221.2	10.1	647	14	CD690890
17	218.6	10.0	482	14	CB161132
18	218	9.9	472	14	W94278
19	218	9.9	939	13	EX437148
20	212	9.7	367	9	AA356204
21	212	9.7	752	10	BE888674
22	211.8	9.7	263	9	AA471173
23	209	9.5	603	9	AL702012
24	207.4	9.5	721	12	BI669261
25	207.4	9.5	817	12	BI603142
26	206.8	9.4	873	12	BG256677
27	206	9.4	354	9	AA355830
28	204.4	9.3	1016	12	BI835314
29	199.8	9.1	480	14	N31546
30	198.8	9.1	1388	13	BQ226850
31	197	9.0	402	9	AV737596
32	196.4	9.0	611	12	BI596305
33	196	8.9	592	10	AW271353
34	195.6	8.9	996	9	AL521077
35	193.6	8.8	649	12	BI561720
36	192.4	8.8	612	13	BX485700
37	191.8	8.8	1201	9	AL534190
38	179	8.2	450	9	AI830394
39	169	7.7	751	13	BQ618394
40	160.8	7.3	438	9	AA115983
41	159	7.3	612	10	BF058420
42	158	7.2	729	14	CA418306
43	157	7.2	681	12	BG914201
44	157	7.2	850	12	BG030589
45	155.8	7.1	694	12	BG621446

ALIGNMENTS

RESULT 1

AU100150

LOCUS

DEFINITION

AU100150 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
LNG15174 similar to Human interferon-gamma induced protein (IFI 16)
gene, mRNA sequence.

ACCESSION

AU100150

VERSION

AU100150.1

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

300 bp mRNA linear EST 05-APR-2001
AU100150 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
LNG15174 similar to Human interferon-gamma induced protein (IFI 16)
gene, mRNA sequence.
AU100150
AU100150.1 GI:13551279
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 300)
Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese J.,
Hata,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S.,
Okubo,K., Suyama,A. and Sugano,S.
In silico mapping of the 5'-ends of human mRNAs using full-length
enriched and 5'-end enriched cDNA libraries constructed by
Oligo-capping method
Unpublished (2001)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and

Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source
Location/Qualifiers
1..300
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UIG15174"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 11.1%; Score 244; DB 9; Length 300;
Best Local Similarity 99.6%; Pred. No. 3.8e-49;
Matches 255; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1715 ACTCCATTTCTCTGGGG-AAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 1773
Db 2 ACTCCATTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 61
QY 1774 AGTGACTCAACCAAGGCGCTTTTCCCTTGTTATCTTTCAGATCTTCAATTTCTTAGCG 1833
Db 62 AGTGACTCAACCAAGGCGCTTTTCCCTTGTTATCTTTCAGATCTTCAATTTCTTAGCG 121
QY 1834 TTTCTGAGATTACAAATCTCTGCGGTTCCGTTTCTGGGAATTTTACTGATTTATCTCCC 1893
Db 122 TTTCTGAGATTACAAATCTCTGCGGTTCCGTTTCTGGGAATTTTACTGATTTATCTCCC 181
QY 1894 CCCTCACACAAATAGCAATGATTCCTGCATTTCTGGAAGATCTCAAGATCTGGACTACTG 1953
Db 182 CCCTCACACAAATAGCAATGATTCCTGCATTTCTGGAAGATCTCAAGATCTGGACTACTG 241
QY 1954 TTGAATAAATTTCCAG 1969
Db 242 TTGAATAAATTTCCAG 257

RESULT 2

AL702095 480 bp mRNA linear EST 04-SEP-2003
LOCUS DKFP686A15155 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKFP686A15155 5', mRNA sequence.

ACCESSION

AL702095 1 GI:19685451

VERSION

EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

REFERENCE

1 (bases 1 to 480)

AUTHORS

Bloecker, H., Boecker, M., Brandt, P., Mewes, W., Weil, B. and Wiemann, S.

TITLE

EST (Bloecker, H., Boecker, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.)

JOURNAL

Unpublished (1999)

COMMENT

Contact: MIPS
MIPS
Incolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.

No sl sequence available.

This clone (DKFP686A15155) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clones@rzpd.de.

FEATURES

source
Location/Qualifiers
1..480
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

/clone="DKFP686A15155"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="vector: pTriplex2; Site_1: SfIIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match 11.0%; Score 240; DB 9; Length 480;
Best Local Similarity 99.6%; Pred. No. 3.9e-48;
Matches 251; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1719 CATTTTCTCTGGGG-AAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 1777
Db 2 CATTTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 61
QY 1778 ACTCAACCAAGGCGCTTTTCCCTTGTTATCTTTCAGATCTTCAATTTCTTAGCGTTTC 1837
Db 62 ACTCAACCAAGGCGCTTTTCCCTTGTTATCTTTCAGATCTTCAATTTCTTAGCGTTTC 121
QY 1838 TGGAGATTACAAATCTCTGCGGTTCCGTTTCTGGGAATTTTACTGATTTATCTCCCCT 1897
Db 122 TGGAGATTACAAATCTCTGCGGTTCCGTTTCTGGGAATTTTACTGATTTATCTCCCCT 181
QY 1898 CACACAAATAGCAATGATTCCTGCATTTCTGGAAGATCTCAAGATCTGGACTACTGTGA 1957
Db 182 CACACAAATAGCAATGATTCCTGCATTTCTGGAAGATCTCAAGATCTGGACTACTGTGA 241
QY 1958 AAAAATTTCCAG 1969
Db 242 AAAAATTTCCAG 253

RESULT 3

LOCUS

BU927876 656 bp mRNA linear EST 18-OCT-2002
AGENCOURT 10434296 NIH MGC 126 Homo sapiens cDNA clone
IMAGE:6653388 5', mRNA sequence.

ACCESSION

BU927876 1 GI:24116606

VERSION

EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

REFERENCE

1 (bases 1 to 656)

AUTHORS

NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbe-remail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LICM2903 row: f column: 12

High quality sequence stop: 555.

FEATURES

source

Location/Qualifiers
1..656
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6653388"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH MGC 126"
/note="Vector: pDNR-LIB; Site_1: SfiI (ggccattagcc);
Site_2: SfiI (ggccgctcgcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -

4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
 5'-AACGAGTGGTATCAAGGAGGCGGCGG-3' and
 5'-ATTCTAGAGCCGAGCGGCGGACATG-3' (30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-1 kb size fraction (other fractions present in NIH MGC 127 and NIH MGC 128). Library created in the laboratory of T. Uedin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 11.0%; Score 240; DB 13; Length 656;
 Best Local Similarity 99.6%; Pred. No. 4e-48;
 Matches 251; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 1719 CATTTTCTCTGGGG-AATAGCAGAAATAGGAGCAGCAGCACTAGTCACTAACTAAGTG 1777
 4 CATTTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCACTAGTCACTAACTAAGTG 63
 1778 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTCGAGATCTTCTTAGCGTTTC 1837
 64 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTCGAGATCTTCTTAGCGTTTC 123
 1838 TGGAGATTACAACTCTCGGGTTCCTGTTCTGGGAATCTTACTGATTTATCTCCCGCT 1897
 124 TGGAGATTACAACTCTCGGGTTCCTGTTCTGGGAATCTTACTGATTTATCTCCCGCT 183
 1898 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGGACTACTGTGA 1957
 184 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGGACTACTGTGA 243
 1958 AAAAATTTCCAG 1969
 244 AAAAATTTCCAG 255

RESULT 4

CF596894 742 bp mRNA linear EST 26-SEP-2003
 LOCUS AGENCOURT 15667536 NICHDS Hs Ovi Homo sapiens cDNA clone
 DEFINITION IMAGE:30704931 5', mRNA sequence.
 CF596894
 CF596894.1 GI:36353819
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 742)
 NIH-MGC <http://mgi.nci.nih.gov/>
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Gregory F. Erickson, Ph.D.
 cDNA Library Preparation: Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDCM254 row: m column: 04
 High quality sequence stop: 498.
 Location/Qualifiers
 1. .742
 /organism="Homo sapiens"

FEATURES

source

/mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30704931"
 /tissue_type="Ovary"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NICHDS Hs Ovi"
 /note="Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccgtctggcc); Library is oligo-dT primed and directionally cloned. Granulosa lutein cells aspirated from preovulatory follicles of normal cycling women undergoing ovulation induction for infertility due to male factor and normal donors. The cells were from follicles stimulated with Lupron, FSH and hCG. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCGCATATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCGGACATG-3' (30)NN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 2.23 kb (range 1.0-4.5 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 11.0%; Score 240; DB 14; Length 742;
 Best Local Similarity 99.6%; Pred. No. 4e-48;
 Matches 251; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 1719 CATTTTCTCTGGGG-AATAGCAGAAATAGGAGCAGCAGCACTAGTCACTAACTAAGTG 1777
 5 CATTTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCACTAGTCACTAACTAAGTG 64
 1778 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTCGAGATCTTCTTAGCGTTTC 1837
 65 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTCGAGATCTTCTTAGCGTTTC 124
 1838 TGGAGATTACAACTCTCGGGTTCCTGTTCTGGGAATCTTACTGATTTATCTCCCGCT 1897
 125 TGGAGATTACAACTCTCGGGTTCCTGTTCTGGGAATCTTACTGATTTATCTCCCGCT 184
 1898 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGGACTACTGTGA 1957
 185 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGGACTACTGTGA 244
 1958 AAAAATTTCCAG 1969
 245 AAAAATTTCCAG 256

RESULT 5

BM819118 497 bp mRNA linear EST 06-MAR-2002
 LOCUS K-EST0086922 S11SNUI Homo sapiens cDNA clone S11SNUI-28-C10 5',
 DEFINITION mRNA sequence.
 BM819118
 BM819118.1 GI:19175531
 EST.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 497)
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 28 row: C column: 10

REFERENCE 1 (bases 1 to 406)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library Preparation: M.B. Soares Lab Clone Distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES Location/Qualifiers
 source 1..406
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2716186"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Sub3"
 /notes="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; The
 NCI CGAP Sub3 library is a subtracted library derived from
 the NCI CGAP Sub1 library, which is a subtracted library
 derived from B1. B1 constitutes a mixture of 21
 normalized or subtracted NCI CGAP libraries:
 NCI CGAP Co4, NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
 NCI CGAP Kid6, NCI CGAP Kid5, NCI CGAP Kid12,
 NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
 NCI CGAP Br2, NCI CGAP Co8, NCI CGAP CLL1, NCI CGAP Lei2,
 NCI CGAP Brn23, NCI CGAP Lu5, NCI CGAP Lu24,
 NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,
 NCI CGAP Brn25. These 21 libraries were pooled and a
 single-stranded DNA preparation of the resulting mixture
 was used as a tracer in a subtractive hybridization with
 a driver whose composition is detailed below:
 NCI CGAP Kid3 pool 1 LLAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE CloneIDs 1322376-1323911,
 1456008-1456775, 1500552-1502855); NCI CGAP Kid5 pool 1
 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs
 1323912-1325831, 1471368-1472903, 1492104-1493255);
 NCI CGAP Lu5 pool 1 LLAM 3575-3582, 3651-3654 (IMAGE
 CloneIDs 1414920-1417991, 1520904-1522439); NCI CGAP GC4
 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
 CloneIDs 1257096-1258631, 1469064-1470983,
 1475592-1476743); NCI CGAP Pr22 pool 1 LLAM 2457-2459,
 2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
 1101192-1101959, 1217928-1220615); NCI CGAP Co10 pool 1
 LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255,
 1145844-1145351). Subtraction was performed as previously
 described [Bonaldo, Lennon & Soares (1996): Normalization
 and Subtraction: Two Approaches To Facilitate Gene
 Discovery. Genome Research 6, 791-806.
 TAG_TISSUE=Lymph node
 TAG_LIB=NCI CGAP_Lym2
 TAG_SEQ=AAATG"

ORIGIN
 Query Match 10.8%; Score 237.4; DB 10; Length 406;
 Best Local Similarity 98.8%; Pred. No. 1.7e-47;
 Matches 238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1729 GGGGAATAGCAGATAGGAGCCAGCCAGCAGCTAGTCAGCTAACTAGTCACTCAACCAAG 1788
 Db 16 GGGCAATAGCAGATAGGAGCCAGCCAGCAGCTAGTCAGCTAACTAGTCACTCAACCAAG 75
 QY 1789 GCGTTTTTCTTGTATCTTTGCGATATCTTCATTTTCTTAGCGTTCTCGAGATTACA 1848
 Db 76 GCGTTTTTCTTGTATCTTTGCGATATCTTCATTTTCTTAGCGTTCTCGAGATTACA 135

QY 1849 ACATCTCGGGTTCGGTTCCTGGAACTTTACTGATTATCTCCCTCACAACAATAA 1908
 Db 136 ACATCTCGGGTTCGGTTCCTGGAACTTTACTGATTATCTCCCTCACAACAATAA 195
 QY 1909 GCATTGATTCCTGCATTTCTGGAAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCA 1968
 Db 196 GCATTGATTCCTGCATTTCTGGAAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCA 255
 QY 1969 G 1969
 Db 256 G 256

RESULT 8
 CD104534
 LOCUS AGENCOURT 13979535 NIH.MGC 186 Homo sapiens cDNA clone
 DEFINITION IMAGE:30371137 5', mRNA sequence.
 ACCESSION CD104534
 VERSION CD104534.1 GI:30757708
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 746)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDCM161 row: e column: 02
 High quality sequence stop: 561.

FEATURES Location/Qualifiers
 source 1..746
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30371137"
 /lab_host="DH10B (TI phage-resistant)"
 /clone_lib="NIH MGC 186"
 /note="Organ: Pooled Skin; Vector: pDNR-LTB; Site 1: SfiI
 (ggccattatggcc); Site 2: SfiI (ggccgcttcggcc); Library is
 oligo-dT primed and directionally cloned. cDNA was
 prepared from a pooled samples of tissues from skin,
 meninges, duramater, pia matter and choroid plexus. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCATTTATGGC-3' and 3' adaptor
 sequence: 5'-ATTCAGAGCGCGGCGCCGACATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 Library"

ORIGIN
 Query Match 10.7%; Score 234; DB 14; Length 746;
 Best Local Similarity 100.0%; Pred. No. 1.2e-46;
 Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1736 AGCAGATAGGAGCAAGCCAGCAGCTAGTCAGCTAACTAGTCACTCAACCAAGGCCTTT 1795
 Db 3 AGCAGATAGGAGCAAGCCAGCAGCTAGTCAGCTAACTAGTCACTCAACCAAGGCCTTT 62

Mon Mar 29 09:55:03 2004

1796 TTCTTGTTATCTTTCAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 1855
 63 TTCTTGTTATCTTTCAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 122
 1856 GCGTTCCGTTTCTGGAACTTTACTGATTTACTCTCCCTCCCTCACACAAATAAGCATTGA 1915
 123 GCGTTCCGTTTCTGGAACTTTACTGATTTACTCTCCCTCCCTCACACAAATAAGCATTGA 182
 1916 TTCTGCAATTTCTGAAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 1969
 183 TTCTGCAATTTCTGAAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 236
 RESULT 9
 AL7111311 608 bp mRNA linear EST 04-SEP-2003
 LOCUS DXF2p686P1779 r1.686 (synonym: hicc3) Homo sapiens cDNA clone
 DEFINITION DXF2p686P1779 5', mRNA sequence.
 ACCESSION AL7111311 GI:19694666
 VERSION AL7111311
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 608)
 AUTHORS Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and
 Wiemann, S.
 TITLE EST (Duesterhoeft, et al.)
 JOURNAL Unpublished (1999)
 COMMENT Contact: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DXF2p686P1779) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 FEATURES
 source
 1. .608
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DXF2p686P1779"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="686 (synonym: hicc3)"
 /note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI; B;
 cDNA-collection"
 ORIGIN
 Query Match 10.6%; Score 233; DB 9; Length 608;
 Best Local Similarity 100.0%; Pred. No. 2.1e-46;
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1737 GCAGATAGGAGCAAGCCAGCAGTACTAGTCACTAAGTACTCAACCAAGGCCCTTTT 1796
 Db 1 GCAGATAGGAGCAAGCCAGCAGTACTAGTCACTAAGTACTCAACCAAGGCCCTTTT 60
 QY 1797 TCCTTGTTATCTTTCAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 1856
 Db 61 TCCTTGTTATCTTTCAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 120
 QY 1857 CGGTTCCGTTTCTGGAACTTTACTGATTTACTCTCCCTCCCTCACACAAATAAGCATTGAT 1916
 Db 121 CGGTTCCGTTTCTGGAACTTTACTGATTTACTCTCCCTCCCTCACACAAATAAGCATTGAT 180
 QY 1917 TCCTGCAATTTCTGAAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 1969
 Db 183 TTCTGCAATTTCTGAAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 236

181 TCCTGCAATTTCTGAAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 233
 BG253409 854 bp mRNA linear EST 13-FEB-2001
 602363319F1 NTH_MGC_90 Homo sapiens cDNA clone IMAGE:4471651 5',
 mRNA sequence.
 ACCESSION BG253409 GI:12763225
 VERSION BG253409
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 854)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM10291 Row: e Column: 06
 High quality sequence stop: 751.
 Location/Qualifiers
 1. .854
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4471651"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NTH_MGC_90"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 ORIGIN
 Query Match 10.6%; Score 233; DB 12; Length 854;
 Best Local Similarity 100.0%; Pred. No. 2.2e-46;
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1737 GCAGATAGGAGCAAGCCAGCAGTACTAGTCACTAAGTACTCAACCAAGGCCCTTTT 1796
 Db 1 GCAGATAGGAGCAAGCCAGCAGTACTAGTCACTAAGTACTCAACCAAGGCCCTTTT 60
 QY 1797 TCCTTGTTATCTTTCAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 1856
 Db 61 TCCTTGTTATCTTTCAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 120
 QY 1857 CGGTTCCGTTTCTGGAACTTTACTGATTTACTCTCCCTCCCTCACACAAATAAGCATTGAT 1916
 Db 121 CGGTTCCGTTTCTGGAACTTTACTGATTTACTCTCCCTCCCTCACACAAATAAGCATTGAT 180
 QY 1917 TCCTGCAATTTCTGAAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 1969
 Db 181 TCCTGCAATTTCTGAAGATCTCAAGATCTGAGTCTGGACTACTGTTGAAAAAATTTCCAG 233
 RESULT 11
 CD252021 873 bp mRNA linear EST 22-MAY-2003
 LOCUS AGENCOURT 14212004 NIH_MGC_180 Homo sapiens cDNA clone
 DEFINITION IMAGE:30383413 5', mRNA sequence.
 ACCESSION CD252021

Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA Project, Sugiyama, T.; Wakamatsu, A.; Irie, R.;
Umezawa, A.; Fukuma, M.; Kuwakari, S.; Hata, J.; Ishii, S.; Yamamoto, J.;
Isono, Y.; Saito, K.; Nakamura, Y.; Masuho, Y.; Nagai, K.; Isogai, T.
HRI human cDNA project; cDNA library construction & 5'-end one
pass sequencing; Helix Research Institute.

FEATURES

Location/Qualifiers
1..579
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CHONS2001343"
/cell_type="chondrocytes"
/clone_lib="CHONS2"
/note="Vector: pME18SFL3"

ORIGIN

Query Match 10.2%; Score 223; DB 9; Length 579;
Best Local Similarity 99.6%; Pred. No. 6.1e-44;
Matches 234; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1736 AGCAGAAATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTGAAGTCAACCAAGGCCCTTTT 1795
Db 1 AGCAGAAATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTGAAGTCAACCAAGGCCCTTTT 60
QY 1796 TTCTCTGTTATCTTTGAGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAACATCCT 1855
Db 61 TTCTCTGTTATCTTTGAGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAACATCCT 120
QY 1856 GCGGTTCCGTTTCTGGGAACCTTTAC-TGATTATCTCCCGCTCACACAAATAGCATTTG 1914
Db 121 GCGGTTCCGTTTCTGGGAACCTTTACNTGATTTATCTCCCGCTCACACAAATAGCATTTG 180
QY 1915 ATTCTCTGATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 1969
Db 181 ATTCTCTGATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 235

Search completed: March 28, 2004, 03:14:27
Job time : 5734 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 05:37:57 ; Search time 158 Seconds
(without alignments)

7836.223 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2.1/USPRO.spool/US09931733/runat 25032004 155916 9625/app_query.fasta_1.2375
-DB=A.Geneseq 29Jan04 -QMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733 @CNG 1 1 167 @runat 25032004 155916 9625 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A.Geneseq 29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95.5	2.5	203	AAU19374	AAU19374 Human G p
2	88	2.3	128	AAU93147	AAU93147 Human epi
3	87	2.3	147	AAU93349	AAU93349 Human epi
4	87	2.3	322	ABP43964	ABP43964 mRNA for
5	87	2.3	331	ABP43964	ABP43964 mRNA for
6	86.5	2.3	944	ABP39507	ABP39507 Staphyloc
7	85.5	2.2	331	AAU90653	AAU90653 Human mut
8	85.5	2.2	331	ADC22703	ADC22703 Human G p
9	85.5	2.2	456	ADA33128	ADA33128 Acinetoba
10	83	2.2	331	AAU44486	AAU44486 Human GPR

C 11	83	2.2	331	4	AAU78390	Human pro
C 12	83	2.2	331	4	AAU38949	Human pol
C 13	83	2.2	331	6	ABP81884	Human G p
C 14	83	2.2	333	4	AAU82048	Human G p
C 15	83	2.2	340	4	AAU79374	Human pro
16	82.5	2.2	310	4	AAU73008	Human pro
17	82.5	2.2	310	4	AAU72789	Human oif
18	82.5	2.2	310	7	ADD47232	Rat Prote
19	81.5	2.1	582	4	AAU03896	G protein
C 20	81	2.1	34	4	AAU17710	Novel hum
C 21	80.5	2.1	447	6	ABP57694	Saccharop
C 22	80	2.1	260	2	AAU27871	Odorant r
C 23	80	2.1	312	2	AAU27873	Odorant r
C 24	80	2.1	312	4	AAU73009	Olfactory
C 25	80	2.1	331	3	AAU90619	Human G p
C 26	80	2.1	331	7	ADC22555	Human G p
C 27	79.5	2.1	272	3	AAU84944	A human c
C 28	79.5	2.1	272	3	AAU97289	Lipid ass
C 29	79.5	2.1	330	5	ABP28451	Streptoco
C 30	79	2.1	362	7	ADC87555	Human GPC
C 31	78	2.0	110	7	ADB36332	Human imm
C 32	78	2.0	118	7	ADB36331	Human imm
C 33	78	2.0	139	7	ADB36333	Human imm
C 34	78	2.0	212	5	AAU47347	Human Pax
C 35	78	2.0	252	6	ABU57655	Different
C 36	78	2.0	377	2	AAU48693	G-protein
C 37	78	2.0	377	2	AAU02665	G-protein
C 38	78	2.0	409	2	AAU03544	Bacillus
C 39	78	2.0	432	3	AAU05849	Arabidops
C 40	77.5	2.0	314	6	ABM69332	Photornab
C 41	77.5	2.0	475	6	ABU38376	Protein e
C 42	77.5	2.0	532	4	ABG10280	Novel hum
C 43	77.5	2.0	631	4	AAU93865	Human pol
C 44	77	2.0	329	4	ABG16890	Novel hum
C 45	77	2.0	1105	4	AAU40174	Human pol

ALIGNMENTS

RESULT 1

AAU19374
ID AAU19374 standard; protein; 203 AA.

XX

AC AAU19374;

XX 04-DEC-2001 (first entry)

XX Human G protein-coupled receptor nGPCR-2307.

Human; G protein-coupled receptor; nGPCR-x; antiviral; analgesic; cytostatic; cardiac; antidiabetic; anorectic; hypotensive; hypertensive; antiparkinsonian; nootropic; neuroprotective; antidepressant; viral infection; HIV-1; human immunodeficiency virus; HIV-2; pair; cancer; metabolic disease; cardiovascular disease; type 2 diabetes; obesity; anorexia; hypotension; hypertension; myocardial infarction; atherosclerosis; Parkinson's disease; psychosis; neurological disorder; schizophrenia; migraine; major depression; anxiety; mental disorder; manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.

XX Homo sapiens.

XX WO200166751-A2.

XX 13-SEP-2001.

XX 08-MAR-2001; 2001WO-US0007370.

XX 08-MAR-2000; 2000US-0187583P.

XX 08-MAR-2000; 2000US-0187584P.

XX 08-MAR-2000; 2000US-0187637P.

XX 08-MAR-2000; 2000US-0187639P.

XX 08-MAR-2000; 2000US-0187640P.

XX 08-MAR-2000; 2000US-0187707P.

08-MAR-2000; 2000US-0187708P.
 08-MAR-2000; 2000US-0187709P.
 08-MAR-2000; 2000US-0187827P.
 08-MAR-2000; 2000US-0188290P.
 08-MAR-2000; 2000US-0188292P.
 08-MAR-2000; 2000US-0188293P.
 (PHAA) PHARMACIA & UPJOHN CO.
 Vogeli G;
 WPI: 2001-536779/59.
 N-PSDB; AAS30946.
 Isolated nucleic acid molecules encoding G protein-coupled receptors termed nGPCR-x, useful in the treatment and diagnosis of viral infections, cancers and mental disorders (e.g. Parkinson's disease and schizophrenia).
 Claim 31; Page 267; 292pp; English.
 The invention relates to novel isolated nucleic acid molecules encoding G protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides, polypeptides, and modulators may be used in the treatment of diseases and conditions such as infections, such as viral infections caused by HIV-1 (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity, anorexia, hypotension, hypertension, myocardial infarction, atherosclerosis), Parkinson's disease, and psychotic and neurological disorders, including schizophrenia, migraine, major depression, anxiety, mental disorder, manic depression, and dyskinesias, such as Huntington's disease or Tourette's Syndrome and many other diseases and syndromes listed in the specification. nGPCR-x polynucleotides and polypeptides, as well as nGPCR-x modulators, may also be used in diagnostic assays for such diseases or conditions. The present sequence represents a G protein-coupled receptor of the invention
 SQ Sequence 203 AA;
 Alignment Scores:
 Pred. No.: 0.0366 Length: 203
 Score: 95.50 Matches: 34
 Percent Similarity: 36.36% Conservative: 14
 Best Local Similarity: 25.76% Mismatches: 41
 Query Match: 2.50% Indels: 43
 DB: 4 Gaps: 4
 US-09-931-733-1 (1-2191) x AAU19374 (1-203)
 QY 1329 CATGTTGCTCTCTTATATTGAGCATCATTTA----- 1361
 Db 91 HisSerAlaLeuMetLeuPheHisValCysLeuPheThrValProGluLeuCys 110
 QY 1362 -----AGGATATATAGTCTCTTAATG 1382
 Db 111 ValCysLeuLeuLeuLeuTyArgIleHisSerLeuLeuLeuLeuAspLeuPhe 130
 QY 1383 ACTCATGTAAGTCTGCTGGGTAGAGAGTCTTAATGTTGGAATCTCTTCAGTATCG 1442
 Db 131 Thr-----Ser 132
 QY 1443 AATTTTGTAGAAAAAACTATTGTTTATATACATTATGATACAAAGTGTAAAT 1502
 Db 133 AsnThrIleTrpGlyPheSerIleCys-----IleAsnLeuArgGlyThrHisAlaVal 150
 QY 1503 TCGTTACATGATATATGCAATTGCTGAAGTTCAGAGATTCAGTGTCCATCACCGG 1562
 Db 151 PheLeuHisArgTyArgAsnAlaGluTrpArgSerLeuLeuLeuLeuLeuLeuLeu 170
 QY 1563 AAAATGTTAACTGATCCCAT---TAAAGTAAATTTCTATCCCATTTCCCTCACCAC 1619
 Db 171 AsnAsnValHisCysThrHisValAsnSerHisProSerAsnProLeuProProSerHis 190

QY 1620 ATGCTCAGCCCTGCGCAGGAACCTGTTTCATTTCT 1655
 Db 191 ProProSerPheLeuAlaSerAsnValTy-Tyr-Ser 202
 RESULT 2
 AAU99347
 ID AAU99347 standard; protein; 128 AA.
 XX AC AAU99347;
 XX DT 07-OCT-2002 (first entry)
 XX DE Human epididymal secretory protein E3 alpha precursor (CCS3, Q14507).
 XX KW Human; vaccine; gene therapy; cytokine; angiogenin; 1B1E.A;
 KW Biopendium search database; CCS3; Q14507;
 KW epididymal secretory protein E3 alpha precursor; immune response;
 KW cell growth; inflammation; wound healing; embryogenesis; apoptosis;
 KW oncology; cancer; reproductive health-related condition.
 XX OS Homo sapiens.
 XX FN WO200244382-A1.
 XX PD 06-JUN-2002.
 XX PF 28-NOV-2001; 2001WO-GB005245.
 XX PR 28-NOV-2000; 2000GB-00028971.
 XX FA (INPH-) INPHARMATICA LTD.
 XX PI Gutteridge A, Fagan RJ, Phelps CB;
 XX WPI: 2002-537454/57.
 XX DR N-PSDB; ABK87961.
 XX Novel isolated CCS3, CCS9 or CCS3a polypeptides which function as cytokines, and polynucleotides encoding the polypeptides, useful as diagnostic reagents for diagnosing disease related to reproductive health in patient.
 Claim 1; Fig 7; 99pp; English.
 The invention discloses three polypeptides which have structures which suggests they can function as cytokines. The polypeptides were identified using the Biopendium search database tool using the human archetypal cytokine family member, angiogenin (PDB code 1B1E:A) structure. The polypeptides identified comprise the epididymal secretory protein E3 alpha precursor (CCS3, Q14507), the epididymal secretory protein E3 precursor (CCS9, CAC17141.1) and the epididymal secretory protein E3 alpha (CCS3a, CAA53971.2). Cytokines are small secreted messenger proteins that act as potential regulators capable of effecting cellular processes, especially the immune response and cell growth, as well as inflammation, wound healing, embryogenesis and development and apoptosis. Cytokines can be used clinically as regulators of the immune system (e.g. in oncology to promote a response against thyroid cancer). The polypeptides are, therefore, useful in therapy and diagnosis of disease, for manufacturing a medicament for treating a reproductive health-related condition, for treating a disease in a patient (e.g. diseases in which there is an aberrant expression of the natural gene or the activity of the polypeptide, where the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist (or antagonist) and as a cytokine. The polypeptide and nucleic acid molecules are also useful for identifying a compound that is effective in the treatment and/or diagnosis of a disease, for chromosome localisation, tissue localisation and for detecting differences in chromosomal location due to e.g. translocation or inversion. The pharmaceutical compositions comprising the polypeptides and/or polynucleotides are useful either as therapeutic or diagnostic compositions, as vaccines and in gene therapy. The sequence presented is the human epididymal secretory protein E3 alpha precursor (CCS3, Q14507)

```
SQ Sequence 128 AA;
Alignment Scores:
Pred. No.: 0.262 Length: 128
Score: 88.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.30% Indels: 33
DB: Gaps: 5

US-09-931-733-1 (1-2191) x AAU99347 (1-128)
QY 230 TCATCATGAAATTTTCATCTTTGGTGAATATGAGCATTCCTGAGT----- 283
Db 3 SerSerLeuLysile-----TrrgLyileLeuLeuAlaLeuLeu 15
QY 284 -----TGGCGCTATAGGTGACCGCAGTTGGTGTGATGCT----- 319
Db 16 CysileLeuCysArgLeu-----CysValTyrSerAsnAsnilleTyrTrp 30
QY 320 -----AGTCCCAATTTGACACATTACTGTCATATAAAATG 352
Db 31 ArgGluPheileLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
QY 353 ATTATAAGTAAAGGAGAGAAATGTTCTGAAGGTCATAAATCCCAAGGGTCTCTTACC 412
Db 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-Ile-- 69
QY 413 CTTTGACCTCTGAGTGGGTTCCAGTGAGTGCCTTAAT 452
Db 70 -TyrSerLeuTrpPheLysileGlnArgAlaCysilleAsn 82

RESULT 3
AAU99349
ID AAU99349 standard; protein; 147 AA.
XX AC AAU99349;
XX DT 07-OCT-2002 (first entry)
XX DE Human epididymal secretory protein E3 alpha (CCS3a, CAA53971.2).
XX KW Human; vaccine; gene therapy; cytokine; angiogenin; lBIE:A;
XX KW Biopendium search database; CCS3a; CAA53971.2;
XX KW epididymal secretory protein E3 alpha; immune response; cell growth;
XX KW inflammation; wound healing; embryogenesis; apoptosis; oncology; cancer;
XX KW reproductive health-related condition.
XX OS Homo sapiens.
XX PN WO200244382-A1.
XX PD 06-JUN-2002.
XX PF 28-NOV-2001; 2001WO-GB005245.
XX PR 28-NOV-2000; 2000GB-00028971.
XX PA (INPH-) INPHARMATICA LTD.
XX PI Gutteridge A, Fagan RJ, Phelps CB;
XX WP1; 2002-537454/57.
XX N-PSDB; ABK37963.
XX Novel isolated CCS3, CCS9 or CCS3a polypeptides which function as
XX cytokines, and polynucleotides encoding the polypeptides, useful as
XX diagnostic reagents for diagnosing disease related to reproductive health
XX in patient.
XX PS Claim 1; Fig 15; 99pp; English.
XX KW The invention discloses three polypeptides which have structures which
```


KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
vulnerary.

OS Homo sapiens.

PN WO200231111-A2.

XX 18-APR-2002.

XX 11-OCT-2001; 2001WO-US027760.

XX 12-OCT-2000; 2000US-00687527.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-426278/45.

XX N-PSDB; ABQ61208.

XX New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.

XX Claim 20; SEQ ID # 867; 357pp + Sequence Listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnerary, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABP43544-
CC ABP43989 represent polypeptides encoded by polynucleotides of the
CC invention. NOTE: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 322 AA;

Alignment Scores:
Pred. No.: 0.565 Length: 322
Score: 87.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.27% Indels: 23
DB: 5 Gaps: 5

US-09-931-733-1 (1-2191) x ABP43964 (1-322)

QY 230 TCNCTATGAATTCATCTTGTGGTATGAATATGAGCATTGCTCGAGT----- 283

DB 178 SerSerLeuLysIle-----TrpGlyIleLeuLeuAlaLeuLeu 190

QY 284 -----TGCGGCTATAGGTGACCGCAGTTTGTGTATGCT----- 319

DB 191 CysIleLeuCysArgLeu-----CysValTyrSerAsnAsnIleTyrTrp 205

QY 320 -----AGTCCCAATGAGACATTACTGCATATAAATG 352

DB 206 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 225

QY 353 ATTATAAGTAAAGGCAGAAATGTTCTGAAGGTCATACTTCCAAGGGTTCTCTACC 412

DB 226 AspValMetArgGluGluAlaLeuLysGlyLysSerPheHisMetPhe-Ile-- 244

QY 413 CTTTGACCTCTCGATTGGGGTTCGCCAGTGCCTTAAT 452

DB 245 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 257

RESULT 5

ABB84727

ID ABB84727 standard; protein; 331 AA.

XX ABB84727,

XX 16-MAY-2002 (first entry)

XX DNA polymerase III holoenzyme delta subunit related protein SEQ ID NO:12.
XX DNA polymerase III holoenzyme delta subunit; DNA polymerase III; DnaX;
KW holo; holB; antibacterial; tuberculostatic; antileptotic;
KW bacterial infection; DNA replication modulation.

XX Buchnera sp.

XX WO200206532-A1.

XX 24-JAN-2002.

XX 16-JUL-2001; 2001WO-US022395.

XX 14-JUL-2000; 2000US-0218246P.

XX 28-MAR-2001; 2001US-00818780.

XX (REPL-) REPLIDYNE INC.

XX Bullard JJ, Janjic N, Mchenry CS;

XX WPI; 2002-164785/21.

XX N-PSDB; ABL87941.

XX Nucleic acid encoding a DNA polymerase III holoenzyme delta subunit from
PT bacteria, useful for screening agents that modulate the subunit activity
PT which is useful in the treatment of bacterial infections e.g. S. pyogenes
PT and S. aureus.

XX Claim 19; Fig 4II; 500pp; English.

XX The present invention describes nucleic acid sequences encoding a DNA
CC polymerase III holoenzyme delta subunit (I) from bacteria. (I) has
CC antibacterial; tuberculostatic; antileptotic. Methods from the present
CC invention can be used for screening for bacterial DNA polymerase
CC holoenzyme delta subunit proteins and agents that modulate their
CC activity. The agents are useful in the treatment of bacterial infections
CC e.g. S. pyogenes, S. aureus, S. pneumoniae, Mycoplasma, Yersinia,
CC Corynebacterium, Salmonella, Mycobacterium tuberculosis or M. leprae. The
CC invention provides a convenient means of identifying compounds which
CC modulate DNA replication in bacteria and therefore provide antibacterial
CC targets, and which are also useful for amplification of DNA. ABL87935 to
CC ABL88071 and ABB84724 to ABB84816 represent sequences used in the
CC exemplification of the present invention

XX SQ Sequence 331 AA;

Alignment Scores:
Pred. No.: 0.573 Length: 331
Score: 87.00 Matches: 48
Percent Similarity: 40.18% Conservative: 40
Best Local Similarity: 21.92% Mismatches: 77
Query Match: 2.28% Indels: 54
DB: 5 Gaps: 12

US-09-931-733-1 (1-2191) x ABB84727 (1-331)

QY 621 AACCATATGAATACTTTATCAAAAAGGAGTAACTCAAGAAAAGCAAGAGGACTG 562

DB 115 AsnHisLeuSerArgPheIleGlnLysAsnLysSerLeuLysGluPheLysAsn----- 132

QY 561 AATATGGCCATAAACAATCTCTGGGTGTCGATATAATCTTTTACTGGCTTAGGGTACTC 502

```

Db      133  -----TyrAsnIleValSerCys-----PheThrProTyrAsnLeuAsn 145
QY      501  TTT---ACGTCGATACATTGAGTTGGCTTAAGCAGTAGTGTAAGAGAGATTAAGGC 445
Db      146  PheIleAsnTrpIle-----
QY      444  ACTCACTGGGAACCCCAATCCAGAGGTCAGAGAGGTCAAGAGGTAGAGAACCCCTTGGAGCTTATGACCC 385
Db      151  ---LysTyrGluIleGlnGluLysLysLysLysLysLysLysLysLysLysLysLys 163
QY      384  TTCAGAACATTTTCTGCTCTTTACTT-----ATAATCATTTTATATATGCA 340
Db      164  ---LysAlaPhePheLeuLeuCysLysTyrTyrGluGlyAsnThrLeuPheIleTyrLys 182
QY      339  GTAATGCTCAATGGGAGTACATACACAACTGGGTCACTATAGCGGCACTC 280
Db      183  IleuAspMetLeuPheIleIleTrpProAspThrCysIleThrGluLysLysLysLys 202
QY      279  AGGCAATGCTCCAT-----AATTATACCAACAAAGATGAAATTTTCATA 235
Db      203  LysIleIleIleGluPhePheAspValSerProSerTyrTrpIleAsnSerIlePheGln 222
QY      234  GATGAACCAAGAAATGCTTAAATGTGC---ATTTTCCCCCACTTTGTGGCAGTGGCC 178
Db      223  GlyLysThrGluLysSerPheTyrIleLeuAsnIlePhe-----PheLysLysLysTyr 240
QY      177  AACCAATGTAATTTGTCTATCATACAGAAAGTCATTAGTAGAATATACAAAATTCCT 118
Db      241  AsnProLeuLeuValArgSerLeuGlnLysAspLeuLeuGlnLeuIleHisMetLys 260
QY      117  GATAAAGCAAAATGTGCGAATATATATATATATATATATATATATATATATATATAT 58
Db      261  ArgGluLysLys-----IleSerIleTyrValMet-----LeuGluLysTyr 274
QY      57  AGTGACTAATTTATATATATATATATATATATATATATATATATATATATATATATAT 1
Db      275  Asn-----IlePheValThrArgArgLysPhePheIleLysAlaPheAsnLysIle 291

```

RESULT 6
ID ABP39507
XX ABP39507 standard; protein; 944 AA.
XX AC ABP39507;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4352.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX OS antibacterial; gene therapy.
XX OS Staphylococcus epidermidis.
XX PN US6380370-B1.
XX PD 30-APR-2002.
XX PF 13-AUG-1998; 98US-00134001.
XX PR 14-AUG-1997; 97US-0055779P.
XX PR 08-NOV-1997; 97US-0064964P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX DR WPI; 2002-381255/41.
XX DR N-PSDB; ABN92052.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX Disclosure; SEQ ID NO 4352; 267pp; English.
PS

```

XX      ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
SQ      Sequence 944 AA;

```

Alignment Scores:	Length:	944
Pred. No.:	86.50	113
Score:	34.63%	91
Percent Similarity:	19.19%	189
Best Local Similarity:	2.26%	196
Query Match:	5	26
DB:		

US-09-931-733-1 (1-2191) x ABP39507 (1-944)

```

QY      81  TATATATGATATATATATTCGCACATTTTTCCTTTATCAGGAATTTTGTATATTTACTA 140
Db      118  TyrLeuSerIleHisLeuMetThrTyrLeuHisTyrSerSerHisSerHisSerLeuMet 137
QY      141  ATGACT-----TTTCGTGTATGTAGC 161
Db      138  LeuThrHisPheLeuThrArgLeuLeuAsnLeuLysArgLeu-CysLeuIle 157
QY      162  ACAATATACAAATTTGGCCACTGC-----CACAAAGTGGGGGAAATGCACA 209
Db      157  eHisLeuMetThrTyrLeuHisCysProSerArgSerHisLeuLeuMetGln-MetHisC 177
QY      210  TTTTAAAGCATTTCTTCTTCATCATATGAAA-----LeuCysLeuLeuIleHisLeu 241
Db      177  ysLeuMetSerMetTyrLeuLeuSerLeuLysArgLeuTyrLeuLeuSerHisSerMetM 197
QY      242  -----ATTTCAATCTTTGTGG-----TATGAATTTATGGA 269
Db      197  etCysLeuTyrCysLeuThrHisSerHisSerLeuMetTrpMetHisTyrLeuMetTrpM 217
QY      270  GCATTGCTCTGAGTTGCGCGCTATAGTGACCGCAGTTCGTGTGTATGTAGTCCCAAT 329
Db      217  etCysLeuLeu-IleLeuAsnHis-----LeuCysLeuLeuIleHisLeu 231
QY      330  GAGACATTAATCTGCATATAAAATGATTATAAGTAAAGGAGCAAAATGTTCTCAAGGTC 389
Db      232  MetThrTyrLeuHis-----
QY      390  ATAACTCCAGGGTCTCTACCCCTTGACCTTGGAATGGGGTTCCTCAGTAGTCCCTT 449
Db      237  -----CysProSerHisSerHis 242
QY      450  AATATCTCTTTCACCTACTGCTTAAGGCCAACTCAATGATATCCAGCTAAAGAGTACCC 509
Db      243  SerLeuMetLeu-MetHisCysLeuMetSer-MetTyrSerLeuAsnLeuLysArgSerT 262
QY      510  TAAAGCCAGTAAAGATTAT-----CACGGAACCCAGAGATCT 548
Db      262  yrLeuMetIleHisLeuTrpMetCysLeuTyrCysLeuThrHisSerHisSerLeuMetL 282
QY      549  TTATGGCCATATCCAGATCCTCTTGTCTTTTCTTGACTTACTCTCTTTTGTATAAAGT 608
Db      282  euMetHisCysLeuMetMetTyrLeuLeuSerLeuLysCysLeuTyrLeuLeuSerH 302
QY      609  TATTATATGGTTTCAATCCAGATTCAGCTATTACCTTCATGCGCCCTAGTTGGATTA 668
Db      302  iserThrMetTyrLeuLeuPheProSerArgSerHisLeuLeuMetGlnMetHisTyrA 322
QY      669  AAACAATCAGCTTCTTTTGT-----CTCTGTAGATGGCTTGGACAGTACATATTAAACAGTC 725

```


Db 227 LysGluLysSerLysArgIleIleThrLeuLeuValGlnValLeuValCys----- 244
QY 528 TAATCTTTTACTGGCTTAGGTTACTCTTTAGCTGGATACATTGAGTTGGCTTAAGGCA 469
Db 245 -----PheMetProPheHisIleCysPheAlaPheLeu-----MetLeuGly 258
QY 468 GTAGTGAAGAAGGATATTAAAGGCACCTCCTCTGGAAACCCCAATCCAGAGGTCAAAGGTA 409
Db 259 ThrGlyGluAsnSerTyr----- 264
QY 408 GAGAACCTTGGGAAGTTATGACCTTTCAGACATTTTCTGCCCTTTTA-----CTT 358
Db 265 ---AsnProTyrGly-----AlaPheThrThrPheLeuMetAsnLeuSerThrCysLeu 281
QY 357 ATAATCATTTTATATGAGTAATGCTCAATTTGGGACTAGCATACACACAACTGCGGTC 298
Db 282 AspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArgValIleSerValMet 301
QY 297 ACCTATAGGCGGCAACTCAGGACATGCTCCATTAATTCATACACAAAGATGAATT 241
Db 302 LeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArgSerGlySerLeu 320
RESULT 8
ADC22703
ID ADC22703 standard; protein; 331 AA.
XX AC ADC22703;
XX DT 18-DEC-2003 (first entry)
XX DE Human G protein-coupled receptor (GPCR) polypeptide #50.
XX KW Human; G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;
XX KW intracellular-3 region; IC3; receptor.
XX OS Homo sapiens.
XX PN US655339-B1.
XX PD 29-APR-2003.
XX PF 13-OCT-1998; 98US-00170496.
XX PR 14-APR-1997; 97US-00839449.
XX PR 14-APR-1998; 98US-00060188.
XX PR 26-JUN-1998; 98US-0090783P.
XX PR 07-AUG-1998; 98US-0095677P.
XX PA (AREN-) ARENA PHARM INC.
XX PI Liaw CW, Behan DP, Chalmers DT;
XX DR WPI; 2003-742861/70.
XX DR N-PSDB; ADC22702.
XX
PT Creating a constitutively active version of an endogenous human G protein
PT coupled receptor (GPCR) comprises substituting a specific amino acid in
PT the transmembrane-6 region with a different amino acid, and testing for
PT constitutive activity.
XX
PS Example 2; SEQ ID NO 184; 221pp; English.
XX
CC The invention relates to a method for treating a non-endogenous,
CC constitutively active version of an endogenous human G protein-coupled
CC receptor (GPCR) that has a transmembrane-6 (TM6) region and an
CC intracellular-3 (IC3) region, by substituting a specific amino acid in
CC the TM6 region with a different amino acid, and testing for constitutive
CC activity. The method is useful for creating a constitutively active
CC version of an endogenous human GPCR that comprises a transmembrane 6
CC region and an intracellular loop 3 region. The altered human GPCR
CC polypeptides are useful for screening test compounds for identification
CC of inverse agonists or partial agonists of GPCR polypeptides, which may
CC have therapeutic uses. The altered GPCRs may also be used in vivo or in

CC vitro in biological research. A nucleic acid encoding the altered GPCR
CC may be used to create a transgenic animal expressing the altered GPCR.
CC The method allows screening for compounds that modulate the activity of a
CC human G protein-coupled receptor without the need for provision of a
CC ligand for the receptor. This is particularly useful in allowing
CC screening of compounds against orphan receptors for which no ligand is
CC currently known. This sequence represents a human GPCR polypeptide of the
CC invention.
XX SQ Sequence 331 AA;
Alignment Scores:
Pred. No.: 0.89 Length: 331
Score: 85.50 Matches: 31
Percent Similarity: 41.18% Conservative: 18
Best Local Similarity: 26.05% Mismatches: 41
Query Match: 2.24% Indels: 29
DB: 7 Gaps: 6
US-09-931-733-1 (1-2191) x ADC22703 (1-331)
QY 585 CAGAAAACCAAGAGG---ATCTGAATATGGCCATAAATCTCTGGGTGTGTGTA 529
Db 227 LysGluLysSerLysArgIleIleThrLeuLeuValGlnValLeuValCys----- 244
QY 528 TAATCTTTTACTGGCTTAGGTTACTCTTTAGCTGGATACATTGAGTTGGCTTAAGGCA 469
Db 245 -----PheMetProPheHisIleCysPheAlaPheLeu-----MetLeuGly 258
QY 468 GTAGTGAAGAAGGATATTAAAGGCACCTCCTCTGGAAACCCCAATCCAGAGGTCAAAGGTA 409
Db 259 ThrGlyGluAsnSerTyr----- 264
QY 408 GAGAACCTTGGGAAGTTATGACCTTTCAGACATTTTCTGCCCTTTTA-----CTT 358
Db 265 ---AsnProTyrGly-----AlaPheThrThrPheLeuMetAsnLeuSerThrCysLeu 281
QY 357 ATAATCATTTTATATGAGTAATGCTCTCAATTTGGGACTAGCATACACACAACTGCGGTC 298
Db 282 AspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArgValIleSerValMet 301
QY 297 ACCTATAGGCGGCAACTCAGGACATGCTCCATTAATTCATACACAAAGATGAATT 241
Db 302 LeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArgSerGlySerLeu 320
RESULT 9
ADA33128
ID ADA33128 standard; protein; 456 AA.
XX AC ADA33128;
XX DT 20-NOV-2003 (first entry)
XX DE Acinetobacter baumannii protein #289.
XX KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX KW plant biocontrol agent.
XX OS Acinetobacter baumannii.
XX PN US6562958-B1.
XX PD 13-MAY-2003.
XX PF 04-JUN-1999; 99US-00328352.
XX PR 09-JUN-1998; 98US-0088701P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton G, Bush D;
XX DR WPI; 2003-576092/54.

DR N-PSDB; ADA29002.

XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
 PT for diagnosing a bacterial disease, as components of antibacterial
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
 PT plants.

XX Example; SEQ ID NO 4415; 328pp; English.

XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents
 CC for diagnosing a bacterial disease, as components of antibacterial
 CC vaccines, as targets for antibacterial drugs, to detect the presence of
 CC A. baumannii and other Acinetobacter species in a sample, in screening
 CC compounds for the ability to interfere with the A. baumannii life cycle
 CC or to inhibit A. baumannii infection, and as biocontrol agents for
 CC plants. The present sequence represents the amino acid sequence of an A.
 CC baumannii protein.

XX SQ Sequence 456 AA;

Alignment Scores:
 Pred. No.: 1.05 Length: 456
 Score: 85.50 Matches: 63
 Percent Similarity: 39.31% Conservative: 51
 Best Local Similarity: 21.72% Mismatches: 101
 Query Match: 2.24% Indels: 75
 DB: 6 Gaps: 12

US-09-931-733-1 (1-2191) x ADA33128 (1-456)

QY 1440 ATAGCTGAAGAGGTTCCCAACATTAGAACTTCTTACCAGCAGACTTACATGTGAGTCA 1381
 Db 64 IleAlaGluHisAsnSerHisValGlnArg---AlaProAlaSerLeuThr---LysMet 81

QY 1380 TTAGAGACTTATATCCCTTAAATGATGCTCAAAATATAAGAG---AAGCAACATGGCTTC 1324
 Db MetValAlaTyrlleAlaLeuLysGluIleLysAlaGlyLysLeuLysLeuAsnGluVal 101

QY 1323 TTTCTCCCTCTTCTCTGTCGACAGAGTGGAGTGGCAGGAGGAG---ATCTTG 1273
 Db 102 IleThrAlaThrProValValSerValValGlnThrPaspGluSerGlnMetTyrLeuLys 121

QY 1272 GTAGGAGCATCTATTGCTACCAATACTTACTGTCATCTCTGTTGGGAAATGCTTCT 1213
 Db 122 AlaGlyGluGlnIleSerValAspGlnLeuLeuAlaGlyLeuIle---ValMetSer 139

QY 1212 ATGAATCGCTATTGTTTAAATACTACTGTAAGTA-----TTCATC 1171
 Db 140 AlaAsnAspAlaAlaValThrLeuAlaGluLysIleSerGlyAspValProHisPheVal 159

QY 1170 CAATATTAATAACTGACAGCTGGCTTCCAGATCAGAAAGTCATATGATATGTA 1111
 Db 160 GlnArgMetAsnGlnGluAlaGlnAlaLeuGlyMetLysAspThrHisPheSer----- 177

QY 1110 AATCCTCTCTTAAATTTATATAGAGAGTTTAAACACTAGCTGCTCTTGCACCATGACTA 1051
 Db 178 AsnProAla-GlyIleThrMetProAspHisTyrThrAla-----HisAspLe 194

QY 1050 GGTACTTACTCTCAATGATATGATTCACGAATGTCAGATTTCTGCAATTATCTCGGA 991
 Db 134 uSerLeuLeuSerGlnAlaVal-IleHisGlnThrProGluTyrLeuHisTyrSerLysM 214

QY 990 TA----- 989

Db 214 etProSerPheSerTyrAsnGlnArgPheHisAlaThrAsnLeuAlaLeuLysTyrA 234

QY 988 -----AACAACGTCGATAGTGGACTCTCAACA 961
 Db 234 spProSerValAspGlyLeuLysThrGlyTyrThrLysAlaAlaGlyTyrAsnLeuAlaL 254

QY 960 TGACAGACTGAGAGACTGGAATAATCTGGAATAATGCACCTGAACTTCGCGTAGTGAGA 901
 ::::::::::::::

Db 254 euThrAlaSer-----Argp 259

QY 900 GTCGTGTTAATCTCTGCTCAGCAGTGTATTTTCTCTGTTAAGGACACTGGGTGCCCTT 841
 Db 259 roSerPheSerProAsnLeuProGlnArgLeuLeuValIleValLeuGlyThrProS 279

QY 840 TAGCGTCTGTAACCCAGCATGTCAATAAATGGCATCACTAGCA----- 794

Db 279 erAlaValLysArgAlaGluIleAlaAspLysLeuMetAsnLeuAlaTyrAlaTyrThrA 299

QY 793 --GATGATGTGTTTAAACCCAGCAGCATCTGCAAGCAGGATGTGTCATGATGCGAA 736
 Db 299 rgAspGluValIleProGlnLysLeuIleAlaGluLeuProValIleLysSer- 318

QY 735 AGGCCCATAGACTGTTAATATGTAC 710

Db 319 -----ThrLeuLysMetPhe 323

RESULT 10
 AAY44486
 ID AAY44486 standard; protein; 331 AA.
 XX
 AC AAY44486;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Human GPRW receptor polypeptide.
 XX
 KW GPRW receptor polypeptide; purinergic receptor; GPCRW receptor; agonist;
 KW antagonist; antibacterial; antifungal; protozoan infection; diabetes;
 KW viral infection; cancer; urinary retention; osteoporosis; allergy;
 KW myocardial infarction; ulcer; psychotic disorder; neurological disorder;
 KW dyskinesia; chromosome 13q32.
 XX
 OS Homo sapiens.
 XX
 PN WO958568-A1.
 XX
 PD 18-NOV-1999.
 XX
 PF 03-MAY-1999; 99WO-US009665.
 XX
 PR 08-MAY-1998; 98US-00075464.
 XX
 PA (SMIK) SMITHLINE BEECHAM CORP.
 XX
 PI Elshourbagy NA;
 XX
 DR WPI; 2000-116288/10.
 DR N-PSDB; AAZ29837.
 XX
 PT New human GPRW polypeptides and polynucleotides used to identify
 PT agonists, antagonists and inhibitors for use in therapy.
 XX
 PS Claim 1; Page 26; 38pp; English.
 CC
 CC The present sequence is the human GPRW receptor polypeptide, which is a
 CC member of the purinergic receptor family and shows structural homology to
 CC GPCRW receptor. GPRW gene is mapped to human chromosome 13q32. GPRW
 CC nucleotide sequence may be used to diagnose or determine susceptibility
 CC to diseases by detection of mutation in GPRW receptor gene and for
 CC chromosome identification. GPRW polypeptides may be used to identify
 CC agonists, antagonists and to detect diseases associated with
 CC inappropriate GPRW activity or levels. GPRW receptor sequence can be used
 CC to treat bacterial, fungal, protozoan and viral infections, particularly
 CC HIV-1 and HIV-2; pain; obesity; anorexia; bulimia; asthma; Parkinson's
 CC disease; acute heart failure; hypotension; hypertension; angina pectoris;
 CC stroke; benign prostatic hypertrophy; migraine; vomiting; cancer;
 CC diabetes; urinary retention; osteoporosis; myocardial infarction; ulcers;
 CC allergies; psychotic and neurological disorders, including anxiety,
 CC schizophrenia, manic depression, depression, delirium, dementia and
 CC severe mental retardation; and dyskinesias such as Huntington's or Gilles
 CC Dele Tourette's syndrome

XX DT 22-OCT-2001 (first entry)
 XX DE Human polypeptide SEQ ID NO 2094.
 XX
 XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX OS Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 XX 26-DEC-2000; 2000WO-US034263.
 XX 23-DEC-1999; 99US-00471275.
 XX 21-JAN-2000; 2000US-00488725.
 XX 25-APR-2000; 2000US-00552317.
 XX 20-JUN-2000; 2000US-00598042.
 XX 19-JUL-2000; 2000US-00620312.
 XX 03-AUG-2000; 2000US-00653450.
 XX 14-SEP-2000; 2000US-00662191.
 XX 19-OCT-2000; 2000US-00693036.
 XX 29-NOV-2000; 2000US-00727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 XX N-PSDB; AAI58105.
 XX DR
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX Example 3; SEQ ID NO 2094; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 XX SQ Sequence 331 AA;
 Alignment Scores:
 Pred. No.: 1.86 Length: 331
 Score: 83.00 Matches: 29
 Percent Similarity: 39.52% Conservative: 20
 Best Local Similarity: 23.39% Mismatches: 47
 Query Match: 2.18% Indels: 28
 DB: 4 Gaps: 5
 US-09-931-733-1 (1-2191) x AAM38949 (1-331)
 QY 603 ATCAAAAAGGAGTAAGTCAAGAAAAGCAAGAGGATCTGGAATATGGCCATAAACATC 544
 XX

Db 222 LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuLeuValGlnVal 241
 QY 543 TCTGGGTGTCGTGATAATCTTTTACTGGCTTTAGGCTACTCTTTACGTGGATACATTGA 484
 Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255
 QY 483 GTTGGCTTAAGCAGTCAGTGAAGAGGATATTAAGGCACCTCACTGGGAACCCCAATCC 424
 Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264
 QY 423 AGAGGTCAAAGGGTAGAGAACCCCTTGGAGATTATGACCCCTTCAGAACATTTTCTGCCTT 364
 Db 265 -----AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn 276
 QY 363 TTA-----CTTATAATCATTATATGCAATGCTCAATTTGGGACTAGCATAC 313
 Db 277 LeuSerThrCysLeuAspValIleLeuTyrTrileValSerLysGlnPheGlnAlaArg 296
 QY 312 ACACAAACTGCGGTCACTATAGCGGCAACTCAGGACAATGCTCCATAATTCATACCAC 253
 Db 297 ValIleSerValMetLeuTyrA-GAsnTyrLeuArgSerMetArgArgLysSerPheArg 316
 QY 252 AAAGATCAAAAT 241
 Db 317 SerGlySerLeu 320
 RESULT 13
 ABP81884
 ID ABP81884 standard; protein; 331 AA.
 XX AC ABP81884;
 XX DT 04-MAR-2003 (first entry)
 XX DE Human G protein-coupled receptor GPR18 protein SEQ ID NO:253.
 XX KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
 KW G protein-coupled receptor modulator; antibody; immune-related disease;
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
 KW immunological-related cell proliferative disease; autoimmune disease;
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
 KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
 KW ulcer.
 XX OS Homo sapiens.
 XX PN WO200261087-A2.
 XX PD 08-AUG-2002.
 XX PF 19-DEC-2001; 2001WO-US050107.
 XX PR 19-DEC-2000; 2000US-0257144P.
 XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.
 XX PI Burner GC, Roush CL, Brown JP;
 XX WPI; 2003-046718/04.
 XX N-PSDB; ABZ42731.
 XX New isolated antigenic peptides e.g., for G protein-coupled receptors
 PT (GPCR), useful for diagnosing and designing drugs for treating conditions
 PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
 PT autoimmune diseases.
 XX Disclosure; Fig 1; 523pp; English.
 XX

CC The present invention describes antigenic peptides (I) comprising: (a)
 CC any one of 1601 sequences (see ABP2019 to ABP83619) of 12-24 amino
 CC acids. Also described: (1) an assay for the detection of a particular G
 CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
 CC and (2) an isolated antibody having high specificity and high affinity or
 CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in
 CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an
 CC antibody against a particular GPCR, and in the production of specific
 CC antibodies. The peptides and antibodies are also useful for detecting the
 CC presence or absence of corresponding GPCRs. The antigenic peptides for
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for
 CC treating immune-related diseases, growth-related diseases, cell
 CC regeneration-related disease, immunological-related cell proliferative
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
 CC any other disorder in which GPCRs are involved. The antibodies may be
 CC used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode
 CC GPCR proteins given in ABP1675 to ABP82018, which are used in the
 CC exemplification of the present invention

XX Sequence 331 AA;

Alignment Scores:
 Pred. No.: 1.86 Length: 331
 Score: 83.00 Matches: 29
 Percent Similarity: 39.52% Conservative: 20
 Best Local Similarity: 23.39% Mismatches: 47
 Query Match: 2.18% Indels: 28
 DB: 6 Gaps: 5

US-09-931-733-1 (1-2191) x ABP81884 (1-331)

QY	603	ATCAAAAAGGAGTAAAGTCAAGAAAAGCAAGAGGAGTCTGGAAATATGGCCATAAATC	544
DB	222	LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuLeuValGlnVal	241
QY	543	TCGTGGGTGTCGTGATAATCTTTTACTGGCTTTAGGCTACTCTTTACGTGGATACATTGA	484
DB	242	LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu-----	255
QY	483	GTGGGCTTAAAGGAGTAAAGTCAAGAAAAGCAAGAGGAGTCTGGAAATATGGCCATAAATC	424
DB	256	-----MetLeuGlyThrGlyGluAsnSerTyr-----	264
QY	423	AGAGGTCAAAGGAGTAAAGGAGTAAAGTCAAGAAAAGCAAGAGGAGTCTGGAAATATGGCCATAAATC	364
DB	265	-----AsnProTyrGly-----AlaPheThrThrPheLeuMetAsn	276
QY	363	TTA-----CTTATAATCATTTTATATGCAAGTAAATGCTCAATTTGGGACTAGCATAC	313
DB	277	LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg	296
QY	312	ACACAAACTCGGTACCTTATAGCGGCAACTCAGGACAAATGCTCCATATATTCATACCAC	253
DB	297	ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg	316
QY	252	AAAGATGAAATT	241
DB	317	SerGlySerLeu	320

RESULT 14
 AAB82048
 ID AAB82048 standard; protein; 333 AA.
 AC AAB82048;
 XX
 XX
 DT 22-JUN-2001 (first entry)

XX Human G protein coupled receptor protein isomer, hGPR-iso.

XX Human; hGPR-iso protein; dendrite cell;

KW G protein coupled receptor protein isomer.

XX Homo sapiens.

OS

XX

FN CN1278006-A.

XX

PD 27-DEC-2000.

XX

PF 30-MAY-2000; 2000CN-00116229.

XX

PR 30-MAY-2000; 2000CN-00116229.

XX

PA (SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.

XX

PI Li N, Xu X, Xiao H;

XX

DR WPI; 2001-245686/26.

DR N-PSDB; AAF84393.

XX

PT Human G protein coupled receptor protein isomer, and coding series

PT thereof.

XX

PS Claim 4; Page 12 (Disclosure); 23pp; Chinese.

XX

CC The present sequence is human G protein coupled receptor protein isomer (hGPR-iso protein), which is expressed in human dendrite cells. The present invention also relates to a preparation method of said protein and nucleic acid sequence and a method of detecting human hGPR-iso nucleic acid sequence and polypeptide in sample

XX Sequence 333 AA;

Alignment Scores:
 Pred. No.: 1.86 Length: 333
 Score: 83.00 Matches: 29
 Percent Similarity: 39.52% Conservative: 20
 Best Local Similarity: 23.39% Mismatches: 47
 Query Match: 2.18% Indels: 28
 DB: 4 Gaps: 5

US-09-931-733-1 (1-2191) x AAB82048 (1-333)

QY	603	ATCAAAAAGGAGTAAAGTCAAGAAAAGCAAGAGGAGTCTGGAAATATGGCCATAAATC	544
DB	224	LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuLeuValGlnVal	243
QY	543	TCGTGGGTGTCGTGATAATCTTTTACTGGCTTTAGGCTACTCTTTACGTGGATACATTGA	484
DB	244	LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu-----	257
QY	483	GTGGGCTTAAAGGAGTAAAGTCAAGAAAAGCAAGAGGAGTCTGGAAATATGGCCATAAATC	424
DB	258	-----MetLeuGlyThrGlyGluAsnSerTyr-----	266
QY	423	AGAGGTCAAAGGAGTAAAGGAGTAAAGTCAAGAAAAGCAAGAGGAGTCTGGAAATATGGCCATAAATC	364
DB	267	-----AsnProTyrGly-----AlaPheThrThrPheLeuMetAsn	278
QY	363	TTA-----CTTATAATCATTTTATATGCAAGTAAATGCTCAATTTGGGACTAGCATAC	313
DB	279	LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg	298
QY	312	ACACAAACTCGGTACCTTATAGCGGCAACTCAGGACAAATGCTCCATATATTCATACCAC	253
DB	299	ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg	318
QY	252	AAAGATGAAATT	241
DB	319	SerGlySerLeu	322

[illegible]

Search completed: March 28, 2004, 08:05:09
Job time : 172 secs

RESULT 15	
AAAM79374	
ID	AAAM79374 standard; protein; 340 AA.
XX	
XX	
AC	AAAM79374;
XX	
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human protein SEQ ID NO 3020.
XX	
KW	Human; cytochrome; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	various system disorder; arthritis; inflammation.

XX	Homo sapiens.
OS	
XX	
XX	WO200157190-A2.
PN	
XX	
XX	09-AUG-2001.
PD	
XX	
XX	05-FEB-2001; 2001WO-US004098.
PF	
XX	
XX	03-FEB-2000; 2000US-00496914.
PR	
PR	27-APR-2000; 2000US-00560875.
XX	
XX	20-JUN-2000; 2000US-00598075.
PN	
PR	19-JUN-2000; 2000US-00620325.
PR	
PR	01-SEP-2000; 2000US-00654936.
XX	
XX	15-SEP-2000; 2000US-00663561.
PR	
PR	20-OCT-2000; 2000US-00693325.
XX	
XX	30-NOV-2000; 2000US-00728422.
XX	
XX	(HYSE-) HYSEQ INC.
PA	
XX	
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI	
PI	Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI	
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX	
XX	WPI; 2001-476283/51.
DR	
DR	N-PSDB; AAK52507.
XX	
XX	Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT	in diagnosis and gene therapy.
PT	
XX	
XX	Claim 20; Page 231-232; 6221pp; English.
PS	
XX	
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC	encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC	(AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC	sequence listing were missing at the time of publication
XX	
XX	Sequence 340 AA;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 07:59:41 ; Search time 32.5 Seconds
(without alignments)
6960.769 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09931733/runat_25032004_155918_9689/app_query.fasta_1.2375
-DB=Issued Patents AA -QFMT=faстан -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptp -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733 -CGN 1_1_30 @runat_25032004_155918_9689 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/pCTUS.COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	2.3	128	4	US-09-755-665-28
2	87	2.3	147	4	US-09-755-665-32
3	86.5	2.3	944	4	US-09-134-001C-4352
4	85.5	2.2	331	4	US-09-170-496D-184
5	85.5	2.2	456	4	US-09-328-352-4415
6	82.5	2.2	1781	2	US-08-477-451-11
7	81.5	2.1	582	4	US-09-721-870-179
8	80.5	2.1	3169	2	US-08-477-451-6
9	80	2.1	331	4	US-09-170-496D-36
10	79.5	2.1	272	4	US-09-177-419C-2
11	78	2.0	377	1	US-08-118-270-14
12	78	2.0	377	5	PCT-US93-08528-14

13	77.5	2.0	538	4	US-09-252-991A-23060	Sequence 23060, A
14	76	2.0	275	1	US-08-118-270-66	Sequence 66, Appl
15	76	2.0	275	5	PCT-US93-08528-66	Sequence 66, Appl
16	76	2.0	415	4	US-09-743-742B-6	Sequence 6, Appl
17	76	2.0	416	3	US-08-858-876A-4	Sequence 4, Appl
18	76	2.0	416	3	US-09-472-880-4	Sequence 4, Appl
19	75.5	2.0	65	4	US-09-540-236-3666	Sequence 3666, Ap
20	75.5	2.0	434	4	US-09-408-020-46	Sequence 46, Appl
21	75	2.0	288	4	US-09-543-681A-6844	Sequence 6844, Ap
22	74.5	2.0	605	2	US-08-752-307B-8	Sequence 8, Appl
23	74.5	2.0	605	4	US-09-707-802-8	Sequence 8, Appl
24	74.5	2.0	605	4	US-09-991-326-8	Sequence 8, Appl
25	74	1.9	462	3	US-09-036-987A-18	Sequence 18, Appl
26	74	1.9	462	3	US-09-370-700-18	Sequence 18, Appl
27	74	1.9	462	4	US-09-603-207-18	Sequence 18, Appl
28	73.5	1.9	1018	1	US-08-408-093-6	Sequence 6, Appl
29	73.5	1.9	1018	1	US-08-408-420A-6	Sequence 6, Appl
30	73.5	1.9	1018	1	US-08-714-901-6	Sequence 6, Appl
31	73.5	1.9	1018	3	US-08-040-741-6	Sequence 6, Appl
32	73.5	1.9	1075	4	US-09-198-452A-916	Sequence 916, App
33	73.5	1.9	3200	2	US-08-477-451-8	Sequence 8, Appl
34	73	1.9	250	4	US-09-252-991A-24614	Sequence 24614, A
35	73	1.9	434	4	US-09-408-020-14	Sequence 14, Appl
36	72	1.9	131	4	US-09-198-452A-1157	Sequence 1157, Ap
37	72	1.9	487	4	US-09-543-681A-6824	Sequence 6824, Ap
38	72	1.9	1305	4	US-08-864-785-3	Sequence 3, Appl
39	72	1.9	1353	3	US-08-894-173-2	Sequence 2, Appl
40	72	1.9	1353	3	US-09-398-193-2	Sequence 2, Appl
41	72	1.9	1353	4	US-09-473-717-3	Sequence 3, Appl
42	71.5	1.9	315	4	US-09-386-653A-9	Sequence 9, Appl
43	71.5	1.9	438	1	US-07-923-095-2	Sequence 2, Appl
44	71.5	1.9	438	1	US-08-229-511-2	Sequence 2, Appl
45	71.5	1.9	438	1	US-08-314-979-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-755-665-28
; Sequence 26, Application US/09755665
; Patent No. 6600019
; GENERAL INFORMATION:
; APPLICANT: Pravaga, Sudhirdas K.
; APPLICANT: Majumdar, Kumud
; APPLICANT: Tailon, Bruce E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-631
; CURRENT APPLICATION NUMBER: US/09/755,665
; PRIOR APPLICATION NUMBER: 2001-08-14
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-755-665-28

Alignment Scores:
Pred. No.: 0.0671
Score: 88.00
Percent Similarity: 43.62%
Best Local Similarity: 29.79%
Query Match: 2.30%
DB: 4
Length: 128
Matches: 28
Conservative: 13
Mismatches: 20
Indels: 33
Gaps: 5

US-09-931-733-1 (1-2191) x US-09-755-665-28 (1-128)

QY 230 TCATCATGAAATTCATCTTTGGTATGAATTATGAGCATTGCTCTGACT----- 283

```

Db      3 SerSerLeuLysIle-----trpGlyIleLeuLeuAlaLeuLeu 15
Qy      284 -----TCCGCGCTATAGGTGACCGACGTTGTGTGTATGCT----- 319
Db      16 CysIleLeuCysArgLeu-----CysValTyrSerAsnIleTyrTrp 30
Qy      320 -----AGTCCCAATTGACACATTACTGCATATAAAATG 352
Db      31 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
Qy      353 ATTATAAGTAAAGCAGAGAAAATGTTCTGAAGGTCATAACTTCCAAAGGGTCTCTACC 412
Db      51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-Ile-- 69
Qy      413 CTTTGACCTCTGGATTGGGTTCCACGAGTGAGTGCCTTAAT 452
Db      70 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 82

RESULT 2
US-09-755-665-32
; Sequence 32, Application US/09755665
; Patent No. 6600019
; GENERAL INFORMATION:
; APPLICANT: Prayaga, Sudhirdas K.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tallon, Bruce E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-631
; CURRENT APPLICATION NUMBER: US/09/755,665
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/174,724
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 32
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-755-665-32

Alignment Scores:
Pred. No.: 0.0966 Length: 147
Score: 87.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.27% Indels: 33
DB: 4 Gaps: 5

US-09-931-733-1 (1-2191) x US-09-755-665-32 (1-147)
Qy      230 TCATCTATGAATTCATCTTTGTGTATGATGAATATATGACGATTGTCTGAGT----- 283
Db      3 SerSerLeuLysIle-----trpGlyIleLeuLeuAlaLeuLeu 15
Qy      284 -----TCCGCGCTATAGGTGACCGACGTTGTGTGTATGCT----- 319
Db      16 CysIleLeuCysArgLeu-----CysValTyrSerAsnIleTyrTrp 30
Qy      320 -----AGTCCCAATTGACACATTACTGCATATAAAATG 352
Db      31 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
Qy      353 ATTATAAGTAAAGCAGAGAAAATGTTCTGAAGGTCATAACTTCCAAAGGGTCTCTACC 412
Db      51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-Ile-- 69
Qy      413 CTTTGACCTCTGGATTGGGTTCCACGAGTGAGTGCCTTAAT 452
Db      70 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 82

```

```

RESULT 3
US-09-134-001C-4352
; Sequence 4352, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4352
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4352

Alignment Scores:
Pred. No.: 0.305 Length: 944
Score: 86.50 Matches: 113
Percent Similarity: 34.63% Conservative: 91
Best Local Similarity: 19.19% Mismatches: 189
Query Match: 2.26% Indels: 196
DB: 4 Gaps: 26

US-09-931-733-1 (1-2191) x US-09-134-001C-4352 (1-944)
Qy      81 TATATATGATATATATATCGCACATTTTGTCTTATCAGGAATTTTGTATATCTACTA 140
Db      118 TyrLeuSerIleHisLeuMetThrTyrLeuHisTyrSerHisSerHisLeuMet 137
Qy      141 ATGACT-----TTTCTGTATGATAGC 161
Db      138 LeuThrHisPheLeuThrLeuThrArgLeuLeuAsnLeuLysArgLeu-CysLeuIle 157
Qy      162 ACAAAATTACAATTGTTGGCCACTGC-----CACAAAGTGGGGGAAATTCACACA 209
Db      157 eHisLeuMetThrTyrLeuHisCysProSerArgSerHisLeuLeuMetGln-MethIsc 177
Qy      210 TTTTAAAGCATTTCTTGTCTCATCTATGAA----- 241
Db      177 ysLeuMetSerMetTyrLeuLeuSerLeuLysArgLeuTyrLeuSerHisSerMetM 197
Qy      242 -----ATTTTCATCTTTGTGG-----TATGAATTATGGA 269
Db      197 etCysLeuTyrCysLeuThrHisSerHisSerLeuMetTrpMetHisTyrLeuMetTrpM 217
Qy      270 GCATTGCTGTAGTTCGGCTATAGTGACCGCAGTTGTGTGTATGTAGTCCCAATT 329
Db      217 etCysLeuLeu-IleLeuAsnHis-----LeuCysLeuLeuIleHisLeu 231
Qy      330 GAGACATTATGCAATATAAATGATTATAAGTAAAGCAGAAAAATGTTCTGAGCGGTC 389
Db      232 MetThrTyrLeuHis----- 236
Qy      390 ATAATTCCCAAGGGTTCTCTACCTTTGACCTCTGAGATTGGGGTCCAGTGAGTGCCTT 449
Db      237 -----CysProSerHisSerHis 242
Qy      450 AATATCCTTCTTACCTACTGCTTAAAGCCAACTCAATGATATCCAGTAAAGATACCC 509
Db      243 SerLeuMetLeu-MethHisCysLeuMetSer-MetTyrSerLeuAsnLeuLysArgSer 262
Qy      510 TAAAGCAGCAATAAAGATTAT-----CAGCAACACCCAGAGATGT 548
Db      262 yrLeuMetIleHisLeuTrpMetCysLeuTyrCysLeuThrHisSerHisSerLeuMetL 282

```

549	QY	TTATGGCCATATTCAGATCCTCTTTGCTTTTCTTGTAATCTCTCTTTTGTATAAAGT	608
282	Db	euMethHisCysLeuMetLeuMetTyrLeuLeuSerLeuLysCysLeuTyrLeuSerH	302
609	QY	TATTATATGCTTCAAAATCCAGAATTCAGCTATTTCACTTCATGGCCCTAGTTGGATTA	668
302	Db	isSerThrMetTyrLeuLeuPheProSerArgSerHisLeuLeuMetGlnMethHisTyrA	322
669	QY	AAACAATCAGCTTTCTTTTG---CTCTGTAGATGGCCTTGACACAGTACATATTAACAGTC	725
322	Db	rgMetLeuMetTyrLeuLeuSerLeuLysArgLeuTyrLeuMetSerHisLeuThrMetC	342
726	QY	TATGGGCCCTTCCACTCATGACACATCC-----TGCTTCAGATCTGC	770
342	Db	ysTrpHisCysLeuSerHisSerHis-SerLeuMetLeuMethHisCysLeuMetSerMet	361
771	QY	TGCTGGGTAAACACACATCATCTGCTAGTTGATGCCAATTA-----	812
362	Db	TyrSerLeuAsnLeuLysAspSerTyrLeuSerSerHisLeuArgMetTyrLeuLeuPhe	389
813	QY	-----TTTGACATCCTCGGGTTACAGCAGCTAAAGGGCACCCAGTGTCTCTTAACA	863
382	Db	ProSerArgSerHisLeuLeuThrLeuMethHisCysLeuMetSerMetTyrSerLeuAsn	401
864	QY	GAGAAAATACACTGCTGACAGAGGATTA-----	893
402	Db	LeuLysAspSerTyrLeuSerSerHisLeuThrMetTyrLeuHisCysLeuThrHisSer	421
894	QY	-----AACACACTCTTCACACATCCCGAAGTT	920
422	Db	HisSerLeuMetTrpTrpHisCysArgMetSerThrArgLeuLeuAsnLeuLysArgLeu	441
921	QY	CAGTGCATTTTCCAGGATTTTCAAGTCTTCGATGCTGATGTTGAGAGTCCCATATCCA	980
442	Db	--CysLeuSerThrArgLeuMetThrTyrPheHisLeuLeuSerHisSerHisSerLeu	460
981	QY	CGTTGTTTATCCAGGAATAATGCAGAAATCTGAGCATTCGTGAATCTAATCATTTGAGG	1040
461	Db	ThrLeuMethis-----CysArgMetSerThrTyrSer-----LeuAsnLeuLys	475
1041	QY	TAAAGTAACTAGTTCATG-----GTCAAGAGGACAGCTAGT	1076
476	Db	ArgSerTyrLeuMetIleHisLeuTrpMetTyrLeuHisPheProSerArgSerHisSer	495
1077	QY	GTAAACTCCTTCAT-----ATAATTTAAAGAAAGGATTTACAT	1115
496	Db	LeuMetLeuMethHisCysArgMetSerThrArgLeuLeuAsnLeuLys-ArgLeuCysLe	515
1116	QY	ATTCATATGACTTTCGTATCTGAAAGCCAGCCTGTCAGTATTAAATTAATGGATGA	1175
515	Db	u-----SerThrArgLeuMetThrTyr-PheHisLeuLeuLys	527
1176	QY	ATACTTCAGGTAGTATTAAACAATAGCGATTTCATAGAA-----GACATTTCCC	1226
527	Db	erHisSerHisSerLeu---ThrLeuMethHisPheLeuThrLeuThrTyrLeuLeuSerP	546
1227	QY	CAACCCAGGAGATGAGTAAAGTA-----TTGGTAGCAATAGATGCTCCTA	1271
546	Db	roLysaspLeuCysLeuSerIleHisLeuMetThrTyrLeuTyrCysLeuThrHisSerH	566
1272	QY	CCAAGATCTCCTCTGCCACTGC-----CACTCTGTGCACAGAAAGGAGGGG	1319
566	Db	isserLeuThrLeuMethHisCysArgMetSerThrTyrSerLeuAsnLeuLys-----	583
1320	QY	AGAAGAACCATGTGCTTCTCTATATTGAGCATCATTTAAGGGGAATATAAGTCTCTA	1379
584	Db	-----CysLeuTyr-Leu-----	587
1380	QY	ATGACTCACATGAAGTCTGCTGGGTAAAGAAAGTTCTTAATGT-----	1421
588	Db	MetIleHisLeuThrMetCysTrpHis-----CysLeuThrHisSerHisSer	603
1422	QY	-----TGGAACTCCTTT---CAGCTATCGAATTTTGAAGAAAAACATTATTGTGT	1472

```

Db      604 LeuMetTrpMetHisPheLeuThrLeuThrArgLeuLeuAsnLeuLysArgLeuCysLeu 623
QY      1473 TATATACAT 1481
Db      624 MetIleHis 626

RESULT 4
US-09-170-496D-184
; Sequence 184, Application US/09170496D
; Patent No.. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 184
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-184

Alignment Scores:
Pred. No.:      0.231      Length:      331
Score:          85.50      Matches:     31
Percent Similarity: 41.18%      Conservative: 18
Best Local Similarity: 26.05%      Mismatches:  41
Query Match:     2.24%      Indels:      29
DB:              4          Gaps:         6

US-09-931-733-1 (1-2191) x US-09-170-496D-184 (1-331)
QY      585 CAGAAAAAGCAAGAGG---ATCTGGAATATGGCCATAAACATCTCTGGGTGTCTGTGA 529
Db      227 LysGluLysSerLysArgIleIleIleThrLeuLeuValGlnValLeuValCys----- 244
QY      528 TAATCTTTTACTGGCTTTAGGGTACTTTTACGTGGATACATTGAGTTGGCTTTAAGGCA 469
Db      245 -----PheMetProPheHisIleCysPheAlaPheLeu-----MetLeuGly 258
QY      468 GTAGGTCAAGAGATATTAAAGGCATCACTGGGAACCCCAATCCAGAGTCAAGAGGTA 409
Db      259 ThrGlyIleAsnSerTyr-----MetLeuGly 264
QY      408 GAGAACCTTGGGAAGTTATGACCCCTTCAGAACATTTTCTGCCTTTTA-----CTT 358
Db      265 ---AsnProTtpGly-----AlaPheThrThrPheLeuMetAsnLeuSerThrCysLeu 281
QY      357 ATAATCATTTTATATGCGAATATGTCTCAATGGGACTAGCATACACACAACTCGCGTC 298
Db      282 AspValIleLeuTyrTyrlleValSerLysGlnPheGlnAlaArgValIleSerVa.Met 301
QY      297 ACCTATAGGCGGCACTCAGGACAAATGCTCCATAATTCATACCACAAAGATGAAT 241
Db      302 LeuTy-ArgAsnTyrLeuArgSerMetArgLysPheArgSerGlySerLeu 320

```

; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4415
 ; LENGTH: 456
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-4415

Alignment Scores:
 Pred. No.: 0.275 Length: 456
 Score: 85.50 Matches: 63
 Percent Similarity: 39.31% Conservative: 51
 Best Local Similarity: 21.72% Mismatches: 101
 Query Match: 2.24% Indels: 75
 DB: 4 Gaps: 12

US-09-931-733-1 (1-2191) x US-09-328-352-4415 (1-456)

QY 1440 ATAGCTGAAGAGGTTTCAACATTAGAACTTTCTTACCAGCAGACATTACATGTGAGTCA 1381
 Db 64 ileAlaGluHisAsnSerHisValGlnArg---AlaProAlaSerLeuThr---LysMet 81
 QY 1380 TTAGAGACTTATTCCTTAAATGATGCTCAATATAAGAG---AAGCAATCGCTTC 1324
 Db 82 MetValAlaTyrlleAlaLeuLysGluileLysAlaGlyLysLeuLysLeuAsnGluVal 101
 QY 1323 TTTCTCCCTCTTCTCTGTCACAGAGTGGCAGTGGCAGGAGGAG-----ATCTTG 1273
 Db 102 lleThrAlaThrProValValSerValValGlnThrPaspGluSerGlnMetTyrLeuLys 121
 QY 1272 GTAGGAGCATCTATGTCACCAATCTTACTGTCATCTCTGTTGGGAAATGTCITCT 1213
 Db 122 AlaGlyGluGlnIleSerValAspGlnLeuLeuAlaGlyLeuIle---ValMetSer 139
 QY 1212 ATGAATCGCTATTGTTTAAATCTACTCTGAGTA-----TTCTATC 1171
 Db 140 AlaAsnAspAlaAlaValThrLeuAlaGluLysIleSerGlyAspValProHisPheVal 159
 QY 1170 CAATATTATTAATCTGACAAAGCCTGGCTTCCAGATCAGAAAGTCATTATGAATATGTA 1111
 Db 160 GlnArgMetAsnGlnGluAlaGlnAlaLeuGlyMetLysAspThrHisPheSer----- 177
 QY 1110 AATCCTTCTTAAATTTATAGAGGACTTTAACTAGCTGTCTCTTGACCATGACTA 1051
 Db 178 AsnProAla-GlyIleThrMetProAspHisTyrThrThraIa-----HisAspLe 194
 QY 1050 GGTACTTCTCAATGATAGTATGATGATGCTCAGATGCTCAGATTTCTGCAATATTCTCTGA 991
 Db 194 uSerLeuLeuSerGlnAlaVal-IleHisGlnThrProGluTyrLeuHisTyrSerLysM 214
 QY 990 TA----- 989
 Db 214 etProSerPheSerTyrAsnGlnArgPheHisAlaThrAsnLeuAlaLeuLysTyrA 234
 QY 988 -----ACAAACGTGGATGATGAGTGGACTCTCAACA 961
 Db 234 spProSerValAspGlyLeuLysThrGlyTyrThrLysAlaAlaGlyTyrAsnLeuAlaL 254
 QY 960 TCACAGCATCGAAGACTTGAATACTCCTGAAATGCACTGAACTCCGGTAGTGTGAGGA 901
 Db 254 euThrAlaSer-----Argp 259
 QY 900 CTCGTGTTTAACTCTGTCAGCATGTGATTTTCTCTGTTAAGGACACTGGGTCCTT 841
 Db 259 roSerPheSerProAsnLeuProGlnArgArgLeuLeuValIleValLeuGlyThrProS 279
 QY 840 TAGCGTGTCTTAACCCAGGATGTCAATTAATGCACTCAACTAGCA----- 794
 Db 279 erAlaValLysArgAlaGluIleAlaAspLysLeuMetAsnLeuAlaTyrAlaTyrThra 299
 QY 793 --GATGATGGTGTAAACCCAGGACAGATCTGCAACGAGGATGTGTCATGAGTGAA 736
 Db 299 rgAspGluValValIleProGluGlnLysLeuIleAlaGluLeuProValIleLysSer- 318

QY 735 AGGCCCCATAGACTGTGTTAATATGTAC 710
 Db 319 -----ThrLeuLysMetPhe 323

RESULT 6

US-08-477-451-11
 ; Sequence 11, Application US/08477451
 ; Patent No. 5928865
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
 ; NUMBER OF SEQUENCES: 46
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,451
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0335.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 510-601-2708
 ; TELEFAX: 510-655-3542
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1781 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-477-451-11

Alignment Scores:
 Pred. No.: 1.36 Length: 1781
 Score: 82.50 Matches: 75
 Percent Similarity: 32.97% Conservative: 45
 Best Local Similarity: 20.60% Mismatches: 109
 Query Match: 2.16% Indels: 135
 DB: 2 Gaps: 18

US-09-931-733-1 (1-2191) x US-08-477-451-11 (1-1781)

QY 243 TTTTCATCTTTGTGTATGAATTTATGAGCATTGCTCCTGAGTTGGCCTTATAGTGACCG 302
 Db 1041 PheHisLeuLysCys-----PheGlyAsnProArgPheLysAlaValPheArgPheGlu 1057
 QY 303 CAGTTTGTGTGTATG-----CTAGTCCCAATTGACACATTACTG 341
 Db 1058 SerValLeuLeuValHisPheGluLeuAspHisPheValIlePheLeuGluArgAspLeu 1077
 QY 342 CATATAAATGATTAATTAAGTAAAGGCAGAAATAATGTTCTGAAGGTCATACTTCAAG 401
 Db 1078 HisVal----- 1079
 QY 402 GGTTCCTCTACCTTTGACCTCTGATTTGG-----GTT 434
 Db 1080 -----LeuAspPheCysLeuGlyIleAsnPheLeuPhePheHisPheAspPheAspIle 1097
 QY 435 CCCAGTGAAGTGCCTTAATATCTCTTCTTCACTACTGCTTAAAGGCCCACTCAATGTATCC 494
 Db 1098 ProGlnCysThrLysArgLeuAsnThrLeuPro--ArgGlnAlaArgCysLeuA 1117

Db 345 SerMetThrAspGluGlnLysLeuSerPheTrpAsnLysLeuSerAsnLysLeuThrPhe 364
QY 2044 AGTATTGTTATTACAGGTTCATGGATGATGAGGAGTAGGCTAGTATTACAGAAAGTGAA 2103
Db 365 Ser-----GlnGlnAspLysThrValGlnHisProAsnPhcGlyHisArgLysSerAsp 382
QY 2104 ACTGAGTTCCTTCATATGATCCTTCA 2130
Db 383 ThrSerIleCysLeuGluAsnProSer 391
RESULT 8
US-08-477-451-6
; Sequence 6, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335,002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-635-3542
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-451-6
Alignment Scores:
Pred. No.: 3 32 Length: 3169
Score: 80.50 Matches: 44
Percent Similarity: 30.39% Conservative: 18
Best Local Similarity: 21.57% Mismatches: 51
Query Match: 2.10% Indels: 91
DB: 2 Gaps: 9
US-09-931-733-1 (1-2191) x US-08-477-451-6 (1-3169)
QY 530 CACGAACACCCAGAGATGTTATGGCCATATCCAGATCCTCTGCTTT---TCTTGA 586
Db 1344 HisHisCysProHisThrThrMetLysLeuPheGlnValLeuLysHisTyrAspSerPro 1363
QY 587 CTTACT-----CCTTTTGTATAAGTATTATCATGTTTCAA 625
Db 1364 LeuThrGlnLysThrGluArgTyrCysGluSerIleIleAspThrCysAsnTrpAspGln 1383
QY 626 ATCCAGAAATTCAGTATTAC----- 646
Db 1384 CysGlnArgLeuProGlnHisLeuThrPheProTyrArgGlnGluLeuLysAsnTyrPro 1403
QY 647 TTCATGGCCCTAGTTTGGATTAAACAATACAGCTTTCTTTTCTGCTGTAGATGGCCTTGG 706

Db 1404 PheHisGlnValAsnLeuLeuGlnAlaValSer-----LeuValCysArgTrpTrpTrp 1421
QY 707 ACAGTACATATTACAGCTCTATGGCCCTTTCCACTCATGACCACATCTCTGCTTGCAGAT 766
Db 1422 LeuTyr-----LeuValGlyHisGlyGluLeuValLeu---Aap 1433
QY 767 CTGCTGCTGGGTTAAACACACATCATCTGCTAGTTGATGCCATTTATTGACATCCTGGG 826
Db 1434 LeuLeuLeuHisGlnLysIleValPheCysGlnIleMet----- 1446
QY 827 GTTACAGCAGCTAAAGGGCACCCAGTGTCTTAACAGAGAGAAAATACACTCTGTCAGCAG 886
Db 1447 -----ArgGluLysTyrProThrAsnPro 1454
QY 886 -----AGGATT 892
Db 1455 HisGlnAlaLeuHisPhePheLeuLeuAlaPheLeuSerArgThrHisGlnThrArg 1474
QY 887 -----AGGATT 892
Db 1475 ProThrPheAspProPheGlyPheAspSerLeuPheLeuGlnLysIleValTrpGlnIle 1494
QY 893 AAACAGACTCCT-----CACACTACCGGAAGTTAGTCATTTCCAGGATTTTCA 943
Db 1495 SerLysThrProLeuSerLeuTyrHisGlnGlyGlyValValTyrPheProAsnPhcGly 1514
QY 944 AGTCTTCGATGC 955
Db 1515 SerIleGlyCys 1518
RESULT 9
US-09-170-496D-36
; Sequence 36, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 36
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-36
Alignment Scores:
Pred. No.: 1.13 Length: 331
Score: 80.00 Matches: 28
Percent Similarity: 39.52% Conservative: 21
Best Local Similarity: 22.58% Mismatches: 47
Query Match: 2.10% Indels: 28
DB: 4 Gaps: 5
US-09-931-733-1 (1-2191) x US-09-170-496D-36 (1-331)
QY 603 ATCAAAAAGGAGTAGTACGACAAAAAGCAAGAGAGATCTGGAATATGCCATAACATC 544
Db 222 LeuLysProLysValLysGluLysSerLeuArgIleIleLeuThrLeuLeuValGlnVal 241
QY 543 TCTGGGTGTCGTGATAATCTTTTACTGCTTTAGGGTACTCTTTACGTGGATACATGCA 484
Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255
QY 483 GTTGGCCTTAAGCGAGTAGTGTGAAGAGATATTAAAGCACTCACTGGGAACCCCAATCC 424
Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264

QY 423 AGAGTCAAGGTAGACACCCCTTGGAGTATGACCCCTTCAGACACATTTTCTGCCTT 364
 Db |||||
 265 ---AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn 276
 QY 363 TTA-----CTTATATCATTTTATATGAGTAATGCTCAGTAATGGAGTACATAC 313
 Db |||||
 277 LeuSerThrCysLeuAspValIleLeuTyrrIleValSerLysGlnPheGlnAlaArg 296
 QY 312 ACACAACTGGCGTCACTATAGGGCGCACTCAGGCAATGCTCCATAATTCATACAC 253
 Db |||||
 297 ValIleSerValMetLeuTyrrArgAsnTyrrLeuArgSerLeuArgArgLysSerPheArg 316
 QY 252 AAAGTGAATT 241
 Db |||||
 317 SerGlySerLeu 320

RESULT 10

US-09-177-419C-2
 ; Sequence 2, Application US/09177419C
 ; Patent No. 6562609
 ; GENERAL INFORMATION:
 ; APPLICANT: Russell, David W
 ; APPLICANT: Lund, Erik G
 ; TITLE OF INVENTION: Cholesterol 25-Hydroxylase
 ; CURRENT APPLICATION NUMBER: US/09/177,419C
 ; CURRENT FILING DATE: 1998-10-22
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 272
 ; TYPE: PRT
 ; ORGANISM: human
 US-09-177-419C-2

Alignment Scores:
 Pred. No.: 1.18 Length: 272
 Score: 79.50 Matches: 46
 Percent Similarity: 38.46% Conservative: 24
 Best Local Similarity: 25.27% Mismatches: 52
 Query Match: 2.08% Indels: 60
 DB: 4 Gaps: 10

US-09-931-733-1 (1-2191) x US-09-177-419C-2 (1-272)

QY 307 TTGTGTATGCTAGTCCCAATGAGACATTAGTGCATATAAATGATTAAGTAAAG 366
 Db |||||
 61 LeuCysSerTrpValProAlaLeuArgArgTyrrLysIleHisProAspPheSerProSer 80
 QY 367 GCAGAAAA-----TGTCTGAAGGGTCATAACTTCCAAGGTCTCTACCTTTG 417
 Db |||||
 81 AlaGlnGlnLeuLeuProCys-----LeuGlyGlnThrLeuTyrrGlnHis 95
 QY 418 ACCTCTGGATTGGGTTCCCAAGTGAAGTCCCTTAATATCTCTTCCACCTACTGCTTAAG 477
 Db |||||
 96 ValMetPheVal---PheProValThrLeuLeu-HisTrpAlaArgSerProAlaLeuLe 114
 QY 478 GCCAACTCAATGTATCCACGTAAAGAGTACCTAAAGCCAGTAAAGATTATACGAACA 537
 Db |||||
 114 uPro-----HisGluAl 118
 QY 538 CCCAGAGATGTTTATGGCCATATCCAGATCCTCTTTCTTCTTCTGACCTATCTCCTTT 597
 Db |||||
 118 aProGluLeuLeuLeuLeuLeuHisIleLeuPheCys-----LeuLeuLeuPhe 135
 QY 598 TTGTATTAAGTATTTCATATGTTTCAATC----- 628
 Db |||||
 135 eAspMetGluPhePheValThrHisLeuLeuHisHisLysValProTrpLeuTyrrArgTh 155
 QY 629 -----CAGATTCA----- 654
 Db |||||
 155 rPheHisLysValHisHisGlnAsnSerSerPheAlaLeuAlaThrGlnTyrrMetSe 175

QY 555 CTAAGTTTGGATTAAACAATCAGCTTT-----CTTTGCTCTG 693
 Db |||||
 175 rValTrpGluLeuPheSerLeuGlyPhePheAspMetMetAsnValThrLeuLeuGlyCy 195
 QY 694 TAGA-----TGGCCTTGGACAGTACATATTAAACAGTCTATGGGCCCTTTCCACTCATGA 747
 Db |||||
 195 sHisProLeuThrThrLeuThrPheHisValValAsnIleTrp---LeuSerValGluAs 214
 QY 748 CCAC 751
 Db |||||
 214 PHis 215

RESULT 11

US-08-118-270-14
 ; Sequence 14, Application US/08118270
 ; Patent No. 5508384

GENERAL INFORMATION:
 ; APPLICANT: Murphy, Randall B.
 ; APPLICANT: Schuster, David I.
 ; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
 ; NUMBER OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
 ; NUMBER OF SEQUENCES: 348
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/118,270
 ; FILING DATE: 09-SEP-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/943,236
 ; FILING DATE: 10-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Townsend, Kevin G.
 ; REGISTRATION NUMBER: 34,033
 ; REFERENCE/DOCKET NUMBER: MURPHY=2A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; TELEX: 248633
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 377 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-118-270-14

Alignment Scores:
 Pred. No.: 2.17 Length: 377
 Score: 78.00 Matches: 35
 Percent Similarity: 43.93% Conservative: 12
 Best Local Similarity: 32.71% Mismatches: 34
 Query Match: 2.04% Indels: 26
 DB: 1 Gaps: 5

US-09-931-733-1 (1-2191) x US-08-118-270-14 (1-377)

QY 1272 CCAAGATCTCTCTCT-----GCCACTGCCACTCTGTGCACAGGAAAG 1313
 Db |||||
 220 ProProGlyProProArgProAlaAlaAlaThrAlaProLeuAlaAsnGlyArg 239
 QY 1314 GAGGGGAGAAGA-----AGCCATGTTGCTCTCTTATTGAGCATCATTTAAGGAA 1367


```

Db      240 AlaGlyLysArgArgProSerArgLeuValAla-----LeuArgGlu 253
QY      1368 TATAAGTCTCTAATGACTCACATGTAAGTCTGGTAAAGAAAGTCTTAATGTTGGAAC 1427
Db      254 GlnLysAlaLeuLysThrLeuGlyIleMetGlyValPheThrLeu---CysTrp-Le 272
QY      1428 TCCCTTACGATCCAAATTTGTAAGAAAGAAAGTCTTATGTTTATATACATTATGA 1487
Db      272 uProPhePheHisArgGluLeuValProAspArgLeuPheValPhe-PheAsnTrpLeuA 292
QY      1488 GATACAAGTGAATTCGTATCATGATATGATATGATATGATGTAAGTACAGAGATTTCAG 1547
Db      292 rg-----TyrAlaAsnSerAlaPheAsnProIleIleT 303
QY      1548 TGTGCACATCACCCGAA 1564
Db      303 YrCysArgSerProAsp 308

RESULT 12
; Sequence 14, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-08528-14

Alignment Scores:
Pred. No.: 2,17 Length: 377
Score: 78.00 Matches: 35
Percent Similarity: 43.93% Conservative: 12
Best Local Similarity: 32.71% Mismatches: 34
Query Match: 2.04% Indels: 26
DB: 5 Gaps: 5

US-09-931-733-1 (1-2191) x PCT-US93-08528-14 (1-377)
QY      1272 CCAAGATCTCTCTCT-----GCCACTGCCACTCTGTGCACAGAAAG 1313

```

```

Db      220 ProGlyProProArgProAlaAlaAlaAlaThrAlaProLeuAlaAsnG.Yarg 239
QY      1314 GAGGGAGAAGA-----AGCCATGTTGCTTCTCTTATATTGAGCATCAATTTAAGGAA 1367
Db      240 AlaGlyLysArgArgProSerArgLeuValAla-----LeuArgGlu 253
QY      1368 TATAAGTCTCTAATGACTCACATGTAAGTCTGGTAAAGAAAGTCTTAATGTTGGAAC 1427
Db      254 GlnLysAlaLeuLysThrLeuGlyIleMetGlyValPheThrLeu---CysTrp-Le 272
QY      1428 TCCCTTACGATCCAAATTTGTAAGAAAGAAAGTCTTATGTTTATATACATTATGA 1487
Db      272 uProPhePheHisArgGluLeuValProAspArgLeuPheValPhe-PheAsnTrpLeuA 292
QY      1488 GATACAAGTGAATTCGTATCATGATATGATATGATGTAAGTACAGAGATTTCAG 1547
Db      292 rg-----TyrAlaAsnSerAlaPheAsnProIleIleT 303
QY      1548 TGTGCACATCACCCGAA 1564
Db      303 YrCysArgSerProAsp 308

RESULT 13
; Sequence 23060, Application US/09252991A
; Patent No. 5551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23060
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23060

Alignment Scores:
Pred. No.: 3,03 Length: 538
Score: 77.50 Matches: 57
Percent Similarity: 36.33% Conservative: 44
Best Local Similarity: 20.50% Mismatches: 89
Query Match: 2.03% Indels: 89
DB: 4 Gaps: 11

US-09-931-733-1 (1-2191) x US-09-252-991A-23060 (1-538)
QY      104 ATTTTCTTTATCAGGAATTTTGTATATTTCTACTAATGACTTTTCTGTATGATAGCAC 163
Db      129 lPheAlaPheGlyGlyCysAlaLeuPheAlaThrSerTyrTyrValVal----- 145
QY      164 AAATTACAA-----TTGTTGGCCACTGCCACAAAG 193
Db      146 -----GlnArgThrSerGlnAlaArgLeuLeuSerAspThrLeuAlaAlaPheThrPhe 163
QY      194 TGGGGGAAAATGCACATT-----TTTAAGCATTTCTTTGTTCTCA 232
Db      164 TrpGlyTrpGlnAlaValIleValGlyAlaValLeuThrLeuProGlnGlyPheThrThr 183
QY      233 TCTATGAAATTTTCATCTTTGTGATGAATTTGAGGACATTGCTCTGAGTGTCCCTCA 292
Db      184 SerLysGluTyrAlaGluLeuGluTrpProLeu-----AlaIleLeuLeuAlaIleVal 201
QY      293 TAGGTGACCCGACGTTTGTGTATGTGTAGTCCCAATTGAGACATTACTGATATATAAATG 352

```

Db 202 TrpIleThrTyAlaIleValPhe-----PheGlyThr 212
Qy 353 ATTAAAGTAAAGGACAGAAATCTCTGAAGGTCATTAATCCAAAGGTT----- 405
Db 213 IleValLysArgLysValHisIleTyValGlyAsnTrpPheTyGlyAlaPheLe 232
Qy 406 -----CTCTACCTTTGACCTCTGCTGATTGGGTTCCCAAGTAGTGCCCTTAATA 453
Db 233 LeuValThrAlaMetLeuHisIleValAsnHisMetSerLeuProValSerTrpPheLys 252
Qy 454 TCCTTCTTCACTACTGCTTAAGGCCCACTCAATGTATCCAGTAAGAGTACCTTAA 513
Db 253 SerTySerAlaTySerGlyAlaThrAspAlaMetVal----- 265
Qy 514 GCCAGTAAAGATTATACGAACACCCAGAGATGTTATGGCCATATTCCAGATCCTCTT 573
Db 266 -----GlnTrpTrpTyGlyHis-----AsnAlaVal 274
Qy 574 TGCCTTTCTTGACTACTCTCTTTTGATAAAGTTATTCATATGG----- 619
Db 275 GlyPhePheLeuThr-ThrGlyPheLeuGlyMetMetTyTrpPheValProLysGlnAl 294
Qy 620 -----TTTCAAAATCCAGAAATTCAGCTATTCACTTCATGCGCCCTAGTT----- 661
Db 294 aGluArgProValTySerTyArgLeuSerIleValHisPheTrpAlaLeuIleSerLe 314
Qy 662 -----TGGATTAAGAAC 672
Db 314 uTyIleTrpAlaGlyProHisHisLeuHisTyThrAlaLeuProAspTrpAlaGlnSe 334
Qy 673 AATCAGCTTT-----CTTTGCTCTGTAGATGGCTTGGACAGTACATATAACAG 723
Db 334 rLeuGlyMetValMetSerLeuIleLeuAlaProSerTrpGlyGlyMetIleAsnGl 354
Qy 724 TCTATGGCCCTTTCCTACTATGACCATCTCTGTCGAGATCTGCTGCTG 775
Db 354 yMetMetThrLeuSerGlyAlaTrpHisLysLeuArgThrAspProIleLeu 371

RESULT 14

US-08-118-270-66
; Sequence 66, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY=2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528

TELEX: 248633

; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-118-270-66

Alignment Scores:

Pred. No.: 3,26 Length: 275
Score: 76.00 Matches: 37
Percent Similarity: 43.17% Conservative: 23
Best Local Similarity: 26.62% Mismatches: 46
Query Match: 1.99% Indels: 33
DB: 1 Gaps: 7

US-09-931-733-1 (1-2191) x US-08-118-270-66 (1-275)

Qy 524 GATTATCAGCAACACCCAGAGATGTTTATGGCCATATTCAGATCCTCTTTGCT----- 577
Db 27 AspSerHisLeuHisThrProMetTyLeuPheLeuSerAsnLeuSerPheSerAspLeu 46
Qy 578 ---TTTCTTTGACTTACTCTCTTTTTCATAAAGTTATTCATATGTTTCAATCCAGAT 634
Db 47 CysPheSerSerValThr-----MetLeuLysLeuLeuGlnAsnIleGlnSerGlnVal 64
Qy 635 TCAGCTATTCACTTCATGCGCCCTAGTTTGGATTAAACAAATCAGCTTTCTTTGCTCTGT 694
Db 65 ProSerIleSerTyAlaGlyCysLeuTrp-----IlePhePheLeuLeuPhe 81
Qy 695 AGATGGCCCTGGACAGTACATATTAACAGTCTATGGG-----CC 733
Db 82 GlyTyTr-LeuGlyAsnPheLeuLeuValAlaMetAlaTyAspArgTyValAlaIleCy 101
Qy 734 CTTTCCACTCATGACCACCA-----TCCTGCTTGCAGATCTG 769
Db 101 sPheProLeuHisTyThrAsnIleMetSerHisLysLeuCysThrCysLeuLeuVal 121
Qy 770 CTGCTGGGTTTAAACACCAATCATCTGCTAGTTGATGCTATTTTATTTGACATCTGGGTT 829
Db 121 lPheTrpIleMetArgSerSerHisAlaMet-----MetIleThrLe 135
Qy 830 ACAGCAGCTTAAAGGCCAGCCAGTGTCTTAAACAGAGAAAAATPACTGCTGAC 884
Db 135 ulleAlaAlaArg-----LeuSerPheCysGluAsnAsnValLeuLeuAsn 150

RESULT 15

PCT-US93-08528-66
; Sequence 66, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236

Search completed: March 28, 2004, 08:15:04
Job time : 50.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 08:12:12 ; Search time 140.5 Seconds

(without alignments)
8160.868 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence:

1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 2130338

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlib
-O=/cgn2_1/USPO.spool/US0931733/runat_25032004_155919_9779/app_query.fasta_1.2375
-DB=published Applications AA_QPW=fasten -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0931733 @CGN 1.1.155 @runat_25032004_155919_9779
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	88	2.3	128	9	US-09-755-665-28	Sequence 28, Appl
2	88	2.3	128	12	US-10-445-641-2	Sequence 2, Appl
3	87	2.3	147	9	US-09-755-665-32	Sequence 32, Appl
4	87	2.3	147	12	US-10-445-641-6	Sequence 6, Appl
5	87	2.3	331	11	US-09-306-179A-12	Sequence 12, Appl
6	85.5	2.2	331	14	US-10-251-385-184	Sequence 184, App
7	85.5	2.2	357	12	US-10-424-599-267968	Sequence 267968,
8	83	2.2	331	9	US-09-826-508-16	Sequence 16, Appl
9	83	2.2	331	14	US-10-225-695A-253	Sequence 4, Appl
10	83	2.2	331	14	US-10-225-695A-253	Sequence 253, App
11	82.5	2.2	310	9	US-09-771-209-75	Sequence 75, Appl
12	81	2.1	34	9	US-09-764-860-328	Sequence 328, App
13	81	2.1	34	14	US-10-074-095-328	Sequence 328, App
14	81	2.1	34	14	US-10-074-095-328	Sequence 328, App
15	80	2.1	312	9	US-09-771-209-77	Sequence 77, Appl
16	80	2.1	331	14	US-10-251-385-36	Sequence 36, Appl
17	80	2.1	331	15	US-10-411-010-18	Sequence 18, Appl
18	80	2.1	355	12	US-10-425-114-54458	Sequence 54458, A
19	79	2.1	362	14	US-10-017-161-2364	Sequence 2364, Ap
20	79	2.1	362	15	US-10-292-798-2008	Sequence 2008, Ap
21	78.5	2.1	65	12	US-10-424-599-23444	Sequence 23444,
22	78.5	2.1	473	15	US-10-369-493-2432	Sequence 2432, Ap
23	78	2.0	122	12	US-10-424-599-150480	Sequence 150480,
24	78	2.0	252	9	US-09-974-298-193	Sequence 193, App
25	78	2.0	457	12	US-10-425-114-54364	Sequence 54364, A
26	78	2.0	490	12	US-10-425-114-43063	Sequence 43063, A
27	78	2.0	646	12	US-10-425-114-65028	Sequence 65028, A
28	77.5	2.0	475	12	US-10-282-122A-66300	Sequence 66300, A
29	77	2.0	356	12	US-10-425-114-67444	Sequence 67444, A
30	77	2.0	614	12	US-10-425-114-71790	Sequence 71790, A
31	77	2.0	1407	15	US-10-369-493-1528	Sequence 1528, Ap
32	76.5	2.0	277	14	US-10-017-161-2342	Sequence 2342, Ap
33	76.5	2.0	658	14	US-10-156-761-14580	Sequence 14580, A
34	76	2.0	263	14	US-10-017-161-2420	Sequence 2420, Ap
35	76	2.0	263	15	US-10-292-798-2060	Sequence 2060, Ap
36	76	2.0	289	12	US-10-425-114-41818	Sequence 41818, A
37	76	2.0	416	14	US-10-205-219-21	Sequence 21, Appl
38	76	2.0	437	14	US-10-017-161-2204	Sequence 2204, Ap
39	76	2.0	437	15	US-10-292-798-1850	Sequence 1850, Ap
40	76	2.0	446	12	US-10-424-599-215419	Sequence 215419,
41	76	2.0	859	10	US-09-978-522-3	Sequence 3, Appl
42	76	2.0	862	10	US-09-978-522-1	Sequence 1, Appl
43	75.5	2.0	226	14	US-10-017-161-2368	Sequence 2368, Ap
44	75.5	2.0	342	15	US-10-369-493-12383	Sequence 12383, A
45	75.5	2.0	434	13	US-10-027-806-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-09-755-665-28
; Sequence 28, Application US/09755665
; Patent No. US20020107186A1
; GENERAL INFORMATION:
; APPLICANT: Prayaga, Sudhirdas K.
; APPLICANT: Majumder, Kumud
; APPLICANT: Taillon, Bruce E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-631
; CURRENT APPLICATION NUMBER: US/09/755,665
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/174,724
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-755-665-28


```

; GENERAL INFORMATION:
;
; APPLICANT: GUTTERIDGE, ALEX
; APPLICANT: FAGAN, RICHARD JOSEPH
; APPLICANT: PHELPS, CHRISTOPHER BENJAMIN
; TITLE OF INVENTION: CYTOKINE-LIKE PROTEINS
; FILE REFERENCE: 674575-2002
; CURRENT APPLICATION NUMBER: US/10/445,641
; CURRENT FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: PCT/GB01/05245
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: GB 0028971.0
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 6
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CCS3a polypeptide
US-10-445-641-6

Alignment Scores:
Pred. No.:
Score: 0.935
Percent Similarity: 87.00
Best Local Similarity: 43.62%
Query Match: 29.79%
DB: 2.27%
Indels: 12
Gaps: 5
Matches: 28
Conservative: 13
Mismatch: 20
Indels: 33
Gaps: 5

```

Query Match:	2,24%	Indels:	80
DB:	12	Gaps:	9
US-09-931-733-1 (1-2191) x US-10-424-599-267968 (1-357)			
QY	595	AGGAGTAAGTCAAGAAAAGCAAGAGATCTGGAATATGCCATAAACATCTCTGGGTG	536
Db	165	ArgSerArgPheAlaValArgCysGluSerLeuAsnThrGlyValSerPheLeuTrp----	183
QY	535	TTTCGTGATAATCTTTTACTGGCTTTTAGGGT-----	506
Db	184	TrpIleValGlyPheTyr-TrpValValSerGlyGlyAspIleLeuLeuGinAspAlaPro	203
QY	505	ACTCTTTACGTGGATACATTGAGTTGGCTTTAAGCGAGTAGGTGAGAAGGATATTAAAG	446
Db	204	ArgLeuTyr-TrpLeuAlaValAlaPheLeu-----	213
QY	445	CACCTCACTGGGAACCCCAATCCAGAGGTCAAAGGGTAGAAGACCCCTTGGAAAGTTATGACC	386
Db	214	-----Al	214
QY	385	CTTCAGAACATTTTTC-----TGCTTTTA-----	361
Db	214	apheAspValPhePheAlaIlePheCysValValLeuAlaCysLeuIleGlyIleAlaLe	234
QY	360	-----CTTATAATCATTTTATATGCAGTAAATGTCCTCAATGGGACT	320
Db	234	uCysCysLeuProCysIleIleAlaIleLeuTyrAlaValagGlyngluGlyAl	254
QY	319	AGCATACACACAACTGCGGTCACTATAGCGGCAACTCAGGACAAATG-----	271
Db	254	asgGluAlaAspLeuSerMetLeuProTyrTyrPheArgIleLeuSerAspValas	274
QY	270	-----CTCATATAATTCATACCACAA	251
Db	274	plysProSerGlyGlyAlaGlySerMetValProIleuThrSerSerAlaTyrLeuG	294
QY	250	AGATGAA-----ATTTCATAGATGAAGCAAGAAATGCTTAAATATGCAATTTTCCC	197
Db	294	uasnGluArgThrLeuLeuLeuGluAspAlaagLucysIleCysLeuCysSerTyrG	314
QY	196	CCACTTTGTGGCAGTG--GCCAACATTTGCTGTATCATACACAGAAAGTCATT	140
Db	314	uasGlyAlaGluLeuHisAlaLeuProCysAsnHisPheHisSer----SerCysII	333
QY	139	AGTAGAATATACAAAAATTCCTGTATAAGCAAAAAATGTGCGAATAT	94
Db	333	eValIysTrpLeuIleCysMetAsnAlaThrCysProLeuCysIleValTyr	341

Alignment Scores:	
Pred. No.:	3.93
Score:	83.00
Percent Similarity:	39.52%
Length:	331
Matches:	29
Conservative:	20

Query Match:	2.18%	Indels:	28
DB:	14	Gaps:	5
US-09-931-733-1 (1-2191) x US-10-225-567A-253 (1-331)			
QY	603	ATCAMAAGAGAGTAAAGTCAAGAAAAGCAAGAGATCTGGAAATATGGCCATAACATC	544
Db	222	LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuLeuValGlnVal	241
QY	543	TCGGGGTTCGTCGTAATCTTTTACTGGCTTTAGGCTACTCTTTACGTGGATACATTGA	484
Db	242	LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu-----	255
QY	483	GTGGCCCTTAAGCAGTAGGTGAAGAAGGATATTAAGGCACCTCACTGGGAACCCCAATCC	424
Db	256	-----MetLeuGlyThrGlyGluAsnSerTyr-----	264
QY	423	AGAGGTCARAGGCTAGAACCCCTCGAGTATATGACCTTCAGAACATTTTCTGCCTT	364
Db	265	-----AsnProIlePgly-----AlaPheThrThrPheLeuMetAsn	276
QY	363	TTA-----OTTATAATCATTTTATATGCAGTAATCTCAATGGGACTAGCATAC	313
Db	277	LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg	296
QY	312	ACAAACTCGCGTCACTATAGCGGCACTCAGGACAATGCTCCATAATTCAATCCAC	253
Db	297	ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg	316
QY	252	AAAGATGAAATT	241
Db	317	SerGlySerLeu	320
RESULT 11			
US-09-771-209-75			
; Sequence 75, Application US/09771209			
; Patent No. US2002006481/7A1			
GENERAL INFORMATION:			
; APPLICANT: Buck, Linda			
; APPLICANT: Axel, Richard			
; TITLE OF INVENTION: ODORANT RECEPTORS AND USES THEREOF			
; FILE REFERENCE: 0575/38586-B/JPW/ADM			
; CURRENT APPLICATION NUMBER: US/09/771,209			
; CURRENT FILING DATE: 2002-01-26			
; PRIOR APPLICATION NUMBER: US 08/129,079			
; PRIOR FILING DATE: 1993-10-05			
; NUMBER OF SEQ ID NOS: 80			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 75			
; LENGTH: 310			
; TYPE: PRT			
; ORGANISM: Rattus sp. I3			
US-09-771-209-75			
Alignment Scores:			
Pred. No.:	4.39	Length:	310
Score:	82.50	Matches:	56
Percent Similarity:	36.82%	Conservative:	39
Best Local Similarity:	21.71%	Mismatches:	62
Query Match:	2.16%	Indels:	101
DB:	9	Gaps:	12
US-09-931-733-1 (1-2191) x US-09-771-209-75 (1-310)			
QY	515	CCAGTAAAGATATTACGAACACCCAGAGATGTTTATGGCCATATCCAGATCCTCTT	574
Db	16	ProIleProGluGluHisGlnHis-----LeuPheTyrAlaLeuPheLeuValMetTyr	33
QY	575	GCTTTTCTGTACTACTCTTTTGTGATAAGTATTATCATATGTTTCAATCCAG---	631
Db	34	-----LeuThrThrIleLeuGlyAsnLeuLeuIleValLeuValGlnLeu	49
QY	632	AATTCAGCTATTCACTTCATGCCCCCTAGTTGGATTAAACCAATCAGCTTTTCTGCTC	691

US-09-931-733-1 (1-2191) x US-09-764-860-328 (1-34)

QY 1581 GGGTACAGTAAATATTTTCGGGTGATGCGACACTGAAATCTCTGACTCCACCAATATG 1522
 Db 8 GlyTyrAsnValHisTyrSerGlyAspGlyCysThrGluSerProAspPheThrIleVal 27

QY 1521 GCATATATCCATGTACAGAA 1501
 Db 28 GlnTyrIleHisValThrGln 34

RESULT 13

US-10-074-095-328

Sequence 328, Application US/10074095

Publication No. US20030077704A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC008C1

CURRENT FILING DATE: 2002-02-14

PRIOR APPLICATION NUMBER: US/10/074,095

PRIOR FILING DATE: 2002-02-14

PRIOR APPLICATION NUMBER: 09/764,860

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION NUMBER: 60/179,065

PRIOR FILING DATE: 2000-01-31

PRIOR APPLICATION NUMBER: 60/180,628

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: 60/214,886

PRIOR FILING DATE: 2000-06-28

PRIOR APPLICATION NUMBER: 60/217,487

PRIOR FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/225,758

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/220,963

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/217,496

PRIOR FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/225,447

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/218,290

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/225,757

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/226,868

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: 60/216,647

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: 60/225,267

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/216,880

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: 60/225,270

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/251,869

PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/235,834

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: 60/234,274

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: 60/234,223

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: 60/228,924

PRIOR FILING DATE: 2000-08-30

PRIOR APPLICATION NUMBER: 60/224,518

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/236,369

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/224,519

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/220,964

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: 60/241,809

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/249,299

PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: 60/241,785

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/244,617

PRIOR FILING DATE: 2000-11-01

PRIOR APPLICATION NUMBER: 60/225,268

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/236,368

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/251,856

PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/251,868

PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/229,344

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/234,997

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: 60/229,343

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/229,287

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/229,513

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: 60/231,413

PRIOR FILING DATE: 2000-09-08

PRIOR APPLICATION NUMBER: 60/229,509

PRIOR FILING DATE: 2000-09-05

PRIOR APPLICATION NUMBER: 60/236,367

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/237,039

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/237,038

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/236,370

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/236,802

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/237,037

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/237,040

PRIOR FILING DATE: 2000-10-02

PRIOR APPLICATION NUMBER: 60/240,960

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/239,935

PRIOR FILING DATE: 2000-10-13

PRIOR APPLICATION NUMBER: 60/239,937

PRIOR FILING DATE: 2000-10-13

PRIOR APPLICATION NUMBER: 60/241,787

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/246,474

PRIOR FILING DATE: 2000-11-08

PRIOR APPLICATION NUMBER: 60/246,532

PRIOR FILING DATE: 2000-11-08

PRIOR APPLICATION NUMBER: 60/249,216

PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: 60/249,210

PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: 60/226,681

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: 60/225,759

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/225,213

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/227,182

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: 60/225,214

PRIOR FILING DATE: 2000-08-14

PRIOR APPLICATION NUMBER: 60/235,836

Mon Mar 29 09:54:58 2004

```

; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match: 2.13% Indels: 0
DB: Gaps: 0

US-09-931-733-1 (1-2191) x US-10-074-095-328 (1-34)
QY 1581 GGGTACAGTTAAACATTTTCGGGTGATGTGCACACTGAAATCTCTGACTTCACCAATATG 1522
Db 8 GlyTyrAsnValHisTyrSerGlyAspGlyCysThrGluSerProAspPheThrIleVal 27
QY 1521 GCATATATCCATGTAACGAAA 1501
Db 28 GlnTyrIleHisValThrGln 34

RESULT 14
US-10-212-872-328
; Sequence 328, Application US/10212872
; Publication No. US20030215893A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C2
; CURRENT APPLICATION NUMBER: US/10/212,872
; CURRENT FILING DATE: 2002-08-07
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 328
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (?)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-212-872-328

Alignment Scores:
Pred. No.: 2,7 Length: 34
Score: 81.00 Matches: 16
Percent Similarity: 74.07% Conservative: 4
Best Local Similarity: 59.26% Mismatches: 7
Query Match: 2.13% Indels: 0
DB: Gaps: 0

US-09-931-733-1 (1-2191) x US-10-212-872-328 (1-34)
QY 1581 GGGTACAGTTAAACATTTTCGGGTGATGTGCACACTGAAATCTCTGACTTCACCAATATG 1522
Db 8 GlyTyrAsnValHisTyrSerGlyAspGlyCysThrGluSerProAspPheThrIleVal 27
QY 1521 GCATATATCCATGTAACGAAA 1501
Db 28 GlnTyrIleHisValThrGln 34

RESULT 15
US-09-771-209-77
; Sequence 77, Application US/09771209
; Patent No. US20020064817A1
; GENERAL INFORMATION:
; APPLICANT: Buck, Linda
; TITLE OF INVENTION: ODORANT RECEPTORS AND USES THEREOF
; FILE REFERENCE: 0575/38586-B/JPW/ADM
; CURRENT APPLICATION NUMBER: US/09/771,209
; CURRENT FILING DATE: 2002-01-26
; PRIOR APPLICATION NUMBER: US 08/129,079
; PRIOR FILING DATE: 1993-10-05
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 77
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Rattus sp. 18
```

US-09-771-209-77

Alignment Scores:

Pred. No.:	8.79	Length:	312
Score:	80.00	Matches:	39
Percent Similarity:	34.48%	Conservative:	21
Best Local Similarity:	22.41%	Mismatches:	39
Query Match:	2.03%	Indels:	75
DB:	9	Gaps:	7

US-09-931-733-1 (1-2191) x US-09-771-209-77 (1-312)

QY	533	GAACACCCAGAGATGTTATGGCCATATTCAGATCCCTCTTTGCTTTCTTTCTTGACTTACT	592
Db	20	GluHisGlnGlnLeuPhePheAlaLeuPheLeuLeuMetTyr-----LeuThr	35
QY	593	CCTTTGTGATAAGTTATTCATATGGTTTCAAAATCCAG---AATCAGCTATTCACCTTC	649
Db	36	ThrPheLeuGlyAsnLeuLeuValValLeuValGlnLeuAspSerHisLeuHisThr	55
QY	650	ATGGCCCTAGTTGGATTAAACAAATCAGCTTCTTTTGTCTCTGT-----	694
Db	56	ProMetTyrLeuPheLeuSerAsnLeuSerPheSerAspLeuCysPheSerSerValThr	75
QY	694	-----	694
Db	76	MetLeuLysLeuLeuGlnAsnIleGlnSerGlnValProSerIleSerTyrAlaGlyCys	95
QY	695	-----AGATGGCCTTGGACAGTACATATTAAACAGT	724
Db	96	LeuThrGlnIlePhePheLeuLeuPheGlyTyr-LeuGlyAsnPheLeuLeuValAl	115
QY	725	CTATGGG-----CCCTTCCACTCATGACCACA-----	752
Db	115	aMetAlaTyrAspArgTyrValAlaIleCysPheProLeuHisTyrThrAsnIleMetSe	135
QY	753	-----TCCTGCTTGACAGTCTGCTGGGTTAACACCAATCATCTGCTAG	799
Db	135	rHisLysLeuCysThrCysLeuLeuValPheTyrIleMetThrSerSer-----	152
QY	800	TTGATGCCATTATTGACATCCCTGGGGTTACAGCAGCTAAAGGGCACCCCA-----	851
Db	153	-----HisAlaMetMetHisThrLeuLeuAl	161
QY	852	-----GTGTCCTTAACAGAGAAATACACTGCTGAGC	884
Db	161	aAlaArgLeuSerPheCysGluAsnAsnValLeuLeuAsn	174

Search completed: March 28, 2004, 08:34:32

Job time : 149.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: March 28, 2004, 07:46:26 ; Search time 52.5 Seconds
(without alignments)
8028.786 Million cell updates/sec

Title: US-09-931-733-1
Perfect score: 3825
Sequence: 1 tacataatgagagatttta.....tgaggactgaacagagaaa 2191

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q=/cgn2_1/uspto_spool/US0931733/runat_25032004_155917_9663/app_query.fasta_1.2375
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM_ext HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0931733@cgn_1_1_57/runat_25032004_155917_9663 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	2.6	175	2	AB0185
2	88	2.3	128	2	S42795
3	87	2.3	331	2	G84981
4	86.5	2.3	3343	2	T42207
5	82.5	2.2	310	2	E23701
6	82.5	2.2	339	2	A47611
7	82.5	2.2	594	2	H48613
8	81.5	2.1	442	2	S50332
9	81.5	2.1	522	2	S09996
10	80	2.1	312	2	G23701
11	80	2.1	478	2	F96842
12	80	2.1	740	2	B84741
13	79.5	2.1	258	2	I56220
14	79	2.1	250	2	AD1956

15	78	2.0	252	2	A34877
16	78	2.0	504	2	C90634
17	78	2.0	670	2	T50312
18	77.5	2.0	475	2	C83452
19	77.5	2.0	950	2	T09076
20	77	2.0	139	2	S24072
21	77	2.0	399	2	S24880
22	77	2.0	420	2	B97304
23	77	2.0	598	2	S55014
24	77	2.0	1226	2	T45788
25	77	2.0	1407	1	BWBYM1
26	76.5	2.0	227	2	F85433
27	76.5	2.0	276	2	T49123
28	76.5	2.0	335	2	H95152
29	76.5	2.0	928	2	F71541
30	76	2.0	203	2	C64019
31	76	2.0	416	2	S68822
32	76	2.0	504	2	C85485
33	76	2.0	510	2	T11943
34	76	2.0	606	2	S35427
35	75.5	2.0	473	2	A42391
36	75.5	2.0	617	2	T23082
37	75.5	2.0	768	2	T18461
38	75.5	2.0	1018	2	JC4211
39	75	2.0	501	2	T11587
40	75	2.0	604	2	AF0187
41	75	2.0	1336	2	T39978
42	75	2.0	2201	2	AH0095
43	74.5	1.9	202	2	T24134
44	74.5	1.9	242	2	C25192
45	74.5	1.9	242	2	T18490

ALIGNMENTS

RESULT 1

AB0185
Probable membrane protein YPO1518 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0185
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-175 <XUR>
A;Cross-references: GB:AL590842; PIDN:CA90341.1; PID:gl5979560; GSPDB:GN00175
C;Genetics:
A;Gene: YPO1518

Alignment Scores:
Pred. No.: 0.0745 Length: 175
Score: 98.00 Matches: 32
Percent Similarity: 43.86% Conservative: 18
Best Local Similarity: 28.07% Mismatches: 32
Query Match: 2.56% Indels: 32
DB: 2 Gaps: 8

US-09-931-733-1 (1-2191) x AB0185 (1-175)

QY	479	CCAACTCAATGATTCACGCTAAAG-----AGTACCCCTAAAGCCAGTAAAGATTATCAC	532
Db	39	ProProSerCysGluArgLysGlnGlnSerAsnIleLeuPro-----His	54
QY	533	GAACACCCAGAGATGTTTATGCGCATATTCAGATCTCTTTGCTTTTCTGACTTACT	592
Db	55	GlutyrPro-----ValGlyLeuThrLeuLeuCysPheSerGlyLeI----	68

R;McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal
submitted to the EMBL Data Library, February 1997
A;Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A;Reference number: 222073
A;Accession: T42207

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA

A;Residues: 1-3343 <MCA>

A;Cross-references: EMBL:U89653; NID:g2443440; PID:g2443441; PIDN:AA71378.1

A;Experimental source: strain Sprague-Dawley; testes

C;Genetics:

A;Gene: BRCA2

C;Superfamily: breast/ovarian cancer tumor suppressor protein, BRCA2 type

Alignment Scores:
Pred. No.: 1.98 Length: 3343
Score: 86.50 Matches: 49
Percent Similarity: 38.34% Conservative: 25
Best Local Similarity: 25.39% Mismatches: 57
Query Match: 2.27% Indels: 62
DB: 2 Gaps: 9

US-09-931-733-1 (1-2191) x T42207 (1-3343)

```
QY 1482 AATGATATTAACACAAATAAGTTTCTTACAAATTCGATAGCTGAAGAGGATCC 1423
D 1315 AsnAlaTyrLys-----LeuGluAsnSerAspValSerLysSerSer 1328
QY 1422 AACATTAGAACTTCTTACCAGCAGACTTACATGTGAGTCATTAGAGACTTATATCCC 1363
D 1329 ThrSerGlyThr-----ValTyrIleAsn 1336
QY 1362 TTAATGATGCTCAAAVA-----TAAGAGAAGCAACATGGCTTCTTCCCTCCCT 1312
D 1337 LysGlyAspSerAspLeuProPheAlaAlaGluLysGlyAsnIleTyr-----Pro 1353
QY 1311 TTCCTGTGCAGAGTGGCAGTGCAGAGGAGATCTTGTGAGGAGCATCTATTGCTACC 1252
D 1354 GluSerCysThrGlnTyrValArgGluGluAsnAlaGlnIleLysGluSerValSerAsp 1373
QY 1251 AAVACTTTACTGCATCTCCTCGTTGGGGAATGTCT---TCTATGAATCGCTATTGTT 1195
D 1374 LeuThrCysLeuGluValMetLysAlaGluLthrCysHisMetLysSer----- 1390
QY 1194 TAAATACTACCTGAAGTATTCATCCAAATTATTAATCACTGAACGCTGGCTTCCAGA 1135
D 1391 -----SerAspLysGluGlnLeu-Prose 1398
QY 1134 TCAGAAAGTCATTATGATATGTAATCTTCCCTTAAATTTATATGAAGAGTTTAAACAC 1075
D 1398 rAspLysMetGluGlnAsn-----MetLysGluPheAsnIle 1410
QY 1074 TAGCTGTCTCTTGACCATGACTAGGTACTTTACTTCAATGATTAGATTCAACGAATGCT 1015
D 1410 eSerPhe-----GlnThrAlaSerGlyLysAsnIle 1420
QY 1014 CAGATTTCTGCATTATCTCTGGATTAACAAACGCTGGATAGTGCATCTCAACATGACAG 955
D 1420 eArgValSerLysGluSerLeuAsnLysSerValAsnIleLeuAspGlnGlu----- 1437
QY 954 CATCGAAGACTTGAATAATCTCGGAAATGCATCGAAC 918
D 1438 -ThrGluAspLeuThrValThrSerAspSerLeuAsn 1449
```

RESULT 5

E23701

olfactory receptor I3 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 26-Aug-1999

C;Accession: E23701

R;Buck, L.; Axel, R.

Cell 65, 175-187, 1991

A;Title: A novel multigene family may encode odorant receptors: a molecular basis for od

A;Reference number: A23701; MUID:91191556; PMID:1840504

A;Accession: E23701

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-310 <BUC>

A;Cross-references: GB:M64385; NID:g205831; PIDN:AAA41748.1; PID:g205832

C;Superfamily: olfactory receptor OR14

C;Keywords: G protein-coupled receptor; transmembrane protein

Alignment Scores:
Pred. No.: 4.14 Length: 310
Score: 82.50 Matches: 56
Percent Similarity: 36.82% Conservative: 39
Best Local Similarity: 21.71% Mismatches: 62
Query Match: 2.16% Indels: 101
DB: 2 Gaps: 12

US-09-931-733-1 (1-2191) x E23701 (1-310)

```
QY 515 CCAGTAAAGATTATCAGCAACCCAGAGATGTTTATGGCCATATTCAGATCCTCTTT 574
D 16 ProIleProGluGluHisGlnHis-----LeuPheTyrAlaLeuPheLeuValMetTyr 33
QY 575 GCITTTTCTTGACTACTCTCTTTTGTATAAAGTTATTTCATATGTTTCAAATCCIG- 631
D 34 -----LeuThrThrIleLeuGlyAsnLeuIleLeuValValGlnLeu 49
QY 632 AATTACGCTATTACCTTTCATGGCCCTAGTTTGGATTAAACAATCAGCTTCTTTTCTC 691
D 50 AspSerGlnLeuHisThrProMetTyrLeuPheLeuSerAsnLeuSerPheSerAspLeu 69
QY 692 TGT-----AGA 697
D 70 CysPheSerSerValThrMetProLysLeuGlnAsnMetArgSerGlnAspThrSer 89
QY 698 TGGCCTTGG-----ACAGTACATATTAAACAGTCTATCGGCC----- 734
D 90 IleProTyrGlyCysLeuAlaGlnThr-TyrPheMetValPheGlyAspMetGln 109
QY 735 -----TTTCCACTCAT 745
D 109 userPheLeuLeuValAlaMetAlaTyrAspArgTyrValAlaIleCysPheProIleHi 129
QY 746 GACACATCC-----TGCTTCAGATCTGCTGGGTAA 781
D 129 sTyrThrSerIleMetSerProLysLeuCysThrCysLeuValLeuLeuLthrPwleLe 149
QY 782 CACCACATCATCTGCTAGT----- 800
D 149 uThrThrSerHisAlaMetMetHisThrLeuLeuAlaAlaArgLeuSerPheCysGluAs 169
QY 801 -----TGATGCCATTTATTGACATCTCGGGTTACAGCCGTAA 841
D 169 nAsnValValLeuAsnPhePheCysAspLeuPheValLeuLeuLysLeuAlaCysSerAs 189
QY 842 AGGCAC-----CCAGTGTCTTAAACAGAGAAAATACACTGCTG- 881
D 189 pThrTyrIleAsnGluLeuMetIlePheIleMetSerThrLeuLeuIleIleIleIlePro 209
QY 882 -----AGCAGAGATTAAACAGACTCTCTCACACTACCGAAGT 919
D 209 ePheLeuIleValMetSerTyrAlaArgIleIleSerSerIleLeuLysValProSerTh 229
QY 920 TCAGTGCAATTTCCAGGATTTTCAAGTCT-----TCGATGCTGCTCATGTT 964
D 229 rGlnGlyIle-CysLysValPheSerThrCysGlySerHisLeuSerValValSerLeuP 249
QY 965 GAGAGTCCACTATC-----CAGTTTGTATTATCCAGGAATAATGCA 1005
D 249 heTyrGlyThrIleIleGlyLeuTyrLeuCysProAlaGlyAsnAsnSer 265
```

RESULT 6

A47611

QY 1826 TCTTAGCCTTTCTCGAGATTAC---AACATCCTGGTTCGGTTCTCGTTCTCGGAACCTTAC 1880
 Db 176 nLeuPheMetGlySerGluTyrCysGlyAlaTyrGlyTyrArgPheTrpAsnMetTyr 195
 RESULT 7
 H48613
 env polyprotein precursor - myeloblastosis-associated virus (strain MAV-2(O)p9) (fragment)
 C:Species: myeloblastosis-associated virus
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Feb-1998
 C:Accession: H48613
 Virol J, V.; Borroughs, K.; Laessle, F.; Crochet, J.; Dambrine, G.; Smith, R.E.; Perbal
 Virol J, V.; Borroughs, K.; Laessle, F.; Crochet, J.; Dambrine, G.; Smith, R.E.; Perbal
 A:Title: Pathogenic potential of myeloblastosis-associated virus: implication of env pr
 A:Reference number: A48613; MUID:9331743; PMID:8393249
 A:Accession: H48613
 A:Status: Preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-594 <JOL>
 C:Superfamily: type C retrovirus env polyprotein
 Alignment Scores:
 Pred. NO.: 4.47 Length: 594
 Score: 82.50 Matches: 58
 Percent Similarity: 36.15% Conservatives: 36
 Best Local Similarity: 22.31% Mismatches: 86
 Query Match: 2.16% Indels: 80
 DB: 2 Gaps: 14
 US-09-931-733-1 (1-2191) x H48613 (1-594)
 QY 1136 CTGGAAGCCAGCGTTGTCAGTTATTAATTTGGATGATACTCTCAGTAGTATTAA 1195
 Db 61 LeuGluGlnProGly---AsnLeuTrpIleThrTrpAlaAsnArgThrGly----- 76
 QY 1196 ACAATAGCCATTTTCATA-----GAAGACATTTCCCAACACGAGATGCAGT 1243
 Db 77 ---GlnThrAspPheCysLeuSerThrGlnSerAlaThrSerProPheGlnThrCys--- 94
 QY 1244 AAGATTGTTGACGATAGATGCTCTTACCAAGATCTCTCTCCACTGCCACTTTGTG 1303
 Db 95 -----LeuIleGlyIleProSerProIleSerGlu----- 104
 QY 1304 CACAGGAAGAGGGGAGAGACGCCATGTTCTCTTATATTTGAGCATCATTTAAG 1363
 Db 105 -----GlyAspPheLys 108
 QY 1364 GGAATATAAGTCTCTAATGACTCACATGTAACTGCTGGTGAAGAAATTTCTAATGTTG 1423
 Db 109 Gly-----TyrValSerAspThrAsnCysThrThrLeuGlyThrAsp----- 122
 QY 1424 GAATCTCTTCAGTATCGAATTTTGTGAGAAAAAACTTATTGTGTTTATATACATTT 1483
 Db 123 ArgLeuValSerAlaAspPheThrGlyGlyProAspAsnSerThrThrLeuThrTyr 142
 QY 1484 ATGAGATACAAAGTGTAATTTTCGTTACATGGATATATGCCATATTTGGTGAAGTCAGAGAT 1543
 Db 143 ArgLysValSerCys-----LeuLeuLeuLysLeuAsnVal 154
 QY 1544 TCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCATTAAGTAATTTCTCAACCCC 1603
 Db 155 SerMetTrpAspGluProProGluLeu-----GlnLeuLeuGly 167
 QY 1604 CATTTCCCTCCACCATGCTCCAGCCCTTCCAGGAACATGTTTCATTTTCTCTGACTAA 1663
 Db 168 SerGlnSerLeuProAsnIleThrAsnIleAla----- 178
 QY 1664 CAGAAACGAAGCTAAAAACACATGGTGGGAGAGTCTCCACATGTTTTCCT-----ACT 1717
 Db 179 -----GlnIleSerGlyIleThrGlyGlyCysValGlyPheArgProGlnGlyVal 195
 QY 1718 CCATTTTCTCTGGGGAATAGCAGATAGGAGCAAGC-----CAGCACTAGTCA 1765

Db 196 ProTriTyLeuGlyTrpSerArgGlnGluAlaThrArgPheLeuLeuArgHisProSer 215
 QY 1766 GCTAACTAGTACTCAACCAAGCCCTTTTCTTGTATCTTTGCAGATCTTCATTT 1825
 Db 216 PheSerTySerThrGluPro-----PheThrValValThrAlaAspArg-HisAs 232
 QY 1826 TCTTAGGTTTCTGGAGATTAC--AACATCTCGGGTTCGGTTCCTGGGAACCTTTAC 1880
 Db 232 nLeuPheMetGlySerGluTyrcysGlyAlaTyrglyTyraArgPheTrpAsnMetTy 251
 RESULT 8
 S50332
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Katharina tunicata mitochondrion
 C:Species: mitochondrion Katharina tunicata
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002
 C:Accession: S50332
 R:Boore, J.L.; Brown, W.M.
 Genetics 138, 423-443, 1994
 A:Title: Complete DNA sequence of the mitochondrial genome of the black chiton, Katharin
 A:Reference number: S50327; MUID:95129806; PMID:7828825
 A:Accession: S50332
 A:Molecule type: DNA
 A:Residues: 1-442 <BOO>
 A:Cross-references: EMBL:U09810
 C:Genetics:
 A:Gene: ND4
 A:Genome: mitochondrion
 A:Genetic code: SGC4
 A:Start codon: ATA
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:
 Pred. No.: 5.57 Length: 442
 Score: 81.50 Matches: 55
 Percent Similarity: 32.06% Conservative: 29
 Best Local Similarity: 20.99% Mismatches: 57
 Query Match: 121 Indels: 14
 DB: 2 Gaps: 14

US-09-931-733-1 (1-2191) x S50332 (1-442)

QY 194 TGGGGGAAATGCACATTTTAAAGCATTTCTC-----TTTGCT 229
 Db 16 TrpTyrlsMetSerTrpIleSerHisPheMetPhePheMetIleThrLeuAsnPhePhe 35
 QY 230 TCATCTATGAATTTTCATCTTTGTGTATGAATATGGAGC----- 271
 Db 36 SerProMetAsnIleProLysIleAsnTyrglu***PheSerAspAsnPheSerValPro 55
 QY 272 ---ATTGCTCTGAGTTGCGCCTATAGTGCACGCGAGTTTGTGTATGCTAGTCCCAAT 328
 Db 56 LeuMetMetLeuSerCysTrp-----IleThrGlyLeuMetIleMetSerSer----- 71
 QY 329 TGACACATTACTGCATATAAAATGATTATAAGTAAGGAGGAGAAATGTTCTGAAGGGT 388
 Db 72 -----TyrLys-----IleLeuLysAsn 77
 QY 389 CATAACTTCAAGGGTTC----- 406
 Db 78 AsnAsnLeuValSerPhePheLeuLeuAsnValMetIleLeuAsnPheIleIleMet 97
 QY 407 -----TCTACCTTTGACCT 421
 Db 98 ValPheThrGlnLysSerLeuPheSerLeuTyrllePhePheGluAlaSerLeuIlePro 117
 QY 422 -----CTGGATTGGGGTCCCGAGTGCAGTGCCTTAATATCCTTCTTCACCT 466
 Db 118 ThrLeuIleLeuIleLeuMetTrpGlyTyrgln-----Pro 129
 QY 467 ACTGCCTTAAGGCCAATCAATGATATCCACGTAAAGAGTACCCTAAAGCCAGTAAAGAT 526
 Db 130 GluArgLeuGln----- 133

QY 527 TATCAGAAACACCCAGAGATGTTTATGGCCATATTCAGATCCTCTTGTCTTTTCTTGA 586
 Db 134 -----AlaGlyMetTyMetMetIleTyThrIleLeuGlyAlaLeu----- 147
 QY 587 CTTACTCTCTTTTGTATAAGTTATTATCATATGTTTCAAAATCCAGAATTCAGTATTCAC 646
 Db 148 -----ProPheLeuIleAsnIlePhePheIleTySer---HisAsnAlaHisLeuAsn 164
 QY 647 TTCATGGCCCTAGTT-----TGATTT 667
 Db 165 LeuLeuIleMetMetSerLeuProIleMetProTyrglnAlaMetIleSerPheTrp 184
 QY 668 AAAACAATCAGCTTTCTTTTCTCTCTAGATGGCCCTTGCACAGTACATATTAACAGTCTA 727
 Db 185 LeuPheIleIleLeuValPheLeuValLysLeuPro-----IleTySerPheHisLeu 202
 QY 728 TGGGCCCTTTCACCTCAT-----GACCACATCTCTGTTGAGATCTG 769
 Db 203 TrpLeuProLysAlaHisValGluAlaProValAlaGlySerMetIleLeuAlaLeu 222
 QY 770 CTGCTG 775
 Db 223 LeuLeu 224
 RESULT 9
 S09996
 nuclear factor I-A1 - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 08-Dec-2000
 C:Accession: S09996
 R:Rupp, R.A.W.; Kruse, U.; Muthaup, G.; Goebel, U.; Beyreuther, K.; Sippel, A.E.
 Nucleic Acids Res. 18, 2607-2616, 1990
 A:Title: Chicken NF1/TGSCA proteins are encoded by at least three independent genes: NF
 A:Reference number: S09995; MUID:90251434; PMID:2339052
 A:Accession: S09996
 A:Molecule type: mRNA
 A:Residues: 1-522 <RUP>
 A:Cross-references: EMBL:X51486; NID:g63661; PIDN:CAA35853.1; PID:g63662
 C:Superfamily: human nuclear factor I
 C:Keywords: DNA binding; transcription regulation

Alignment Scores:
 Pred. No.: 5.68 Length: 522
 Score: 81.50 Matches: 60
 Percent Similarity: 35.71% Conservative: 35
 Best Local Similarity: 22.56% Mismatches: 92
 Query Match: 2.13% Indels: 79
 DB: 2 Gaps: 13

US-09-931-733-1 (1-2191) x S09996 (1-522)

QY 412 CTTTCCACCTCTGGATTGGG-----TTCCAGTCAGTCCTTAATATCTCTTCTCACC 465
 Db 235 ProIleAlaIaGlyThrGlyProAsnPheSerLeuSerAspLeuGluSerSerTy 254
 QY 466 TACTGCTTAAAGCCCAACTCAATGATATCCAGCTAAAGAGTACCCTAAAGCAGT----- 519
 Db 255 TyrSerMetSerProGlyAlaMet-----ArgArgSerLeuProSerThrSerThr 272
 QY 520 -----AAAAGATTATCAGAACACCCAGAGATGTTTATGGCCATATTCAGATCCT 570
 Db 273 SerSerThrLysArgIleLysSerValGluAspGluMetAspSerProGlyGluGluPro 292
 QY 571 CTTTGGCTTTTCTTGACTTAC-----TCTTTT 597
 Db 293 -----PheTyThrSerGlnGlyArgSerProGlySerGlyserGlnSerSerGlyTrp 310
 QY 598 TTTGATAAAGTTATTATCATATGTTTCAAAATCCAGAATTCAGCTTATTCATTCATGCCCT 657
 Db 311 HisGluValGluProGlyTyrlleAsnProGluHisArgGlyAlaLeuHisGly--- 329
 QY 658 AGTTTGGATTAAACAACATCAGCTTTCTTGTCTGTAGATGGCCTTGGACAGTACATAT 717

```

Db      330  -----MetProSerProThrAlaLeu 336
QY      718  TAACAGTCTATGGCCCTTTCCACTCANGACCAATCCTGCTGCAGATCTGCTGGTGGG 777
Db      337  LysylSerGluylsSerGlyPheSerProSerProSerProSerGlnThrSerSerLeuGly 356
QY      778  -----TTAACACACATCATCTCTGCTAGTTGATCCATTTATTGACATCCTGGGGTTAC 831
Db      357  ThrAlaPheThrGlnHisArgProVal 366
QY      832  AGCAGCGCTAAAGGCGACCCAGTGTCTTAAACAGAGAAAAATACACTGCTGAGCAGAGGAT 891
Db      367  -----lleThrGly-Pro 370
QY      892  TAAACAGACTCCTCAGACTACCGGAAGTTCAGTGCATTTCCAGGATTTCAAGTCCTGCG 951
Db      371  ArgAlaSerProHisAlaThrProSerThrLeuHisPhePro--ThrSerProIle-- 388
QY      952  ATGCTGTCATGTTGAGAGTCCACTATCCAGTGTGTTATCCAGGAATA 1000
Db      389  -----lleGlnProGlyProTyroPheSerHisProAlaIleArgTyroHisPr 405
QY      1001  -ATGACAGAAATCTGAGCATTCGTGAATCTATCATTTGAGGTAAGTAACCTAGTCATGG 1059
Db      405  oGlnGluThrLeuLysGluPheValGlnLeuValCys-----ProAspAlaG 421
QY      1060  TCAAGAGCAGCTAGTGTAACTCCTTCAATATAAATTTAAGGAGGATTTACATATTC 1119
Db      421  yGlnGln-AlaGlyValGlyPheLeuAsnProAsnGlySerSerGinglyLysValH 441
QY      1120  ATAATGACTTTCTG 1133
Db      441  lAsnProPheLeu 445

```

RESULT 10

```

G23701
A:Title: olfactory receptor I8 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 26-Aug-1999
C:Accession: G23701
R:Buck, L.; Axel, R.
Cell 65, 175-187, 1991
A:Title: A novel multigene family may encode odorant receptors: a molecular basis for od
A:Reference number: A23701; MUID:91191556; PMID:1840504
A:Accession: G23701
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-312 <BUC>
A:Cross-references: GB:M64387; NID:g205835; PIDN:AAA41750.1; PID:g205836
C:Superfamily: olfactory receptor OR14
C:Keywords: G protein-coupled receptor; transmembrane protein

```

```

Alignment Scores:
Pred. No.: 7.83 Length: 312
Score: 80.00 Matches: 39
Percent Similarity: 34.48% Conservative: 21
Best Local Similarity: 22.41% Mismatches: 39
Query Match: 2.09% Indels: 75
DB: 2 Gaps: 7

```

US-09-931-733-1 (1-2191) x G23701 (1-312)

```

QY      533  GAACACCCAGAGATCTTATGTCATATCCAGATCCTCTTTGCTTTTCTTGACTTACT 592
Db      20  GluHisGlnGlnPhePheAlaLeuPheLeuIleMetTyro--LeuThr 35
QY      593  CTTTCTTTTATGAAGTGTATTCATATGTTTCCAAATCCAG--AAATCAGCTATTCACCTC 649
Db      36  ThrPheLeuGlyAsnLeuLeuValValLeuValGlnLeuAspSerHisLeuHisThr 55
QY      650  ATGGCCCTAGTTGGATTAACATACATCAGCTTTCTTTGCTCTGT 694

```

```

Db      56  ProMetTyroPheLeuSerAsnLeuSerPheSerAspLeuCysPheSerSerValThr 75
QY      694  ----- 694
Db      76  MetLeuLysLeuLeuGlnAsnIleGlnSerGlnValProSerIleSerTyroAlaGlyCys 95
QY      695  -----AGATGGCCTTGGACAGTACATATTAACAGT 724
Db      96  LeuThrGlnIlePhePhePheLeuLeuPheGlyTyro-LeuGlyAsnPheLeuLeuValAl 115
QY      725  CTATGGG-----CCCTTTCACCTCATGACCACA 752
Db      115  aMetAlaTyroAspArgTyroValAlaIleCysPheProLeuHisTyroThrAsnIleVetSe 135
QY      753  -----TCTGCTTGCAGATCTGCTCTGGTTAAACACACATCATCTGCTAG 799
Db      135  rHisLysLeuCysThrCysLeuLeuLeuValPheThrPheIleMetThrSerSer----- 152
QY      800  TTGATGCCATTTATTGACATCTCTGGGTTACAGCAGCTTAAAGGGCACCCA 851
Db      153  -----HisAlaMetMetHisThrLeuLeuAl 161
QY      852  -----GTGCTCTTAAACAGAGAAAAATACATCTGCTGAGC 884
Db      161  aAlaArgLeuSerPheCysGluAsnAsnValLeuLeuAsn 174

```

RESULT 11

```

F96842
A:Title: hypothetical protein F23A5.32 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F96842
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Klaykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F96842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-478 <STO>
A:Cross-references: GB:AF005173; NID:g6503308; PIDN:AAF14684.1; GSPDB:GN00141
C:Genetics:
A:Gene: F23A5.32
A:Map position: 1

```

```

Alignment Scores:
Pred. No.: 8.23 Length: 478
Score: 80.00 Matches: 46
Percent Similarity: 39.77% Conservative: 24
Best Local Similarity: 26.14% Mismatches: 68
Query Match: 2.09% Indels: 38
DB: 2 Gaps: 10

```

US-09-931-733-1 (1-2191) x F96842 (1-478)

```

QY      1094  AAATTAAAGGAAGGATTTACATATTCATATGACTTTCTGATCTGGAAGCCAGGCTTG 1153
Db      165  LysPheLysThr---PheAsnPheSer-----ProAspSerLeuSerHisProSerLeu 181
QY      1154  TCAGTTATTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1213
Db      182  Met-----SerLeuSerLeuHisSerTyroPheLeu 191
QY      1214  GAACACATTTCCCAACACGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1261
Db      192  GluSerSerHisProLeuArgAsnCysSerAsnLeuArgThrLeuLysLeuLeuSerIle 211

```


QY 559 ATCCAGATCCCTTTGCTTTTCTTGACTTACTCCTTTTTCATTAAGTTATTCATAG 618
 Db |||||
 81 CysProAspPro-----Val 85

QY 619 GTTCAAAATCCAGATTTCAGCTATTCACTTTCATGGCCT 657
 Db ::|||
 86 LeuValAsnGlyGluPheSerSer-----GlyPro 96

Search completed: March 28, 2004, 08:13:47
 Job time : 67.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 05:41:11 ; Search time 29 seconds
(without alignments)
7867.980 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence:

1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2 1/USPTO.spool/US09931733/runat_25032004 155916 9635/app_query.fasta.1.2375
-DB=SwissProt_42 -QFMT=fastan -SURFIX=rsp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733@cgn 1 1 30 @runat 25032004 155916 9635 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	2.3	147	1 EP3A HUMAN	Q14507 homo sapien
2	87	2.3	331	1 HOL4_BUCAI	P57520 buchnera ap
3	86.5	2.2	3343	1 BRC2 RAT	Q35923 rattus norv
4	82.5	2.2	310	1 OLFO RAT	P23269 rattus norv
5	81.5	2.1	522	1 NFIA_CHICK	F17923 gallus gall
6	80	2.1	312	1 OLFO RAT	P23271 rattus norv
7	80	2.1	331	1 GP18 HUMAN	Q14330 homo sapien
8	80	2.1	365	1 GCST_CHLITE	Q8kbj9 chlorobium
9	79.5	2.1	258	1 GRAB RAT	P49864 rattus norv
10	79	2.1	408	1 TRUD_LEPIN	Q8f8n2 leptospira
11	79	2.1	466	1 SHUT_HELVI	Q25190 heliothis v
12	78.5	2.1	473	1 PRPS_SCHPO	Q13615 schizosacch
13	78.5	2.1	653	1 ST12_ARATH	Q9max3 arabidopsis
14	78	2.0	252	1 C4BB HUMAN	P20851 homo sapien
15	78	2.0	409	1 AMB1_BACST	P37113 bacillus st
16	78	2.0	409	1 DXR_SYNPX	Q7u8c3 synchococc
17	78	2.0	504	1 CAIT_ECO57	Q8xa30 escherichia
18	77	2.0	399	1 BR53_CAVPO	P35371 cavia porce

19	77	2.0	598	1 NUSM_PETMA	Q35543 petromycon
20	77	2.0	1407	1 T2D2_YEAST	P32555 saccharomyc
21	76.5	2.0	643	1 ZP64_MOUSE	Q99ke8 mus musculu
22	76	2.0	203	1 NYFI_HAEIN	P44102 haenophilus
23	76	2.0	416	1 NTR2 RAT	O63384 rattus norv
24	76	2.0	457	1 NU4M_ARBLI	P34941 arbutia lix
25	75.5	2.0	434	1 GSA_CERSY	O74038 cenarchaeum
26	75	2.0	501	1 ALG8_SCHPO	Q10479 schizosacch
27	75	2.0	1336	1 NAMI_SCHPO	P78966 schizosacch
28	74.5	1.9	242	1 CRP1_LIMPO	P06205 limulus pol
29	74.5	2.0	456	1 OST4_HUMAN	P39656 homo sapien
30	74.5	1.9	765	1 PHVA_ANASP	Q31sc2 anabaena sp
31	74.5	2.0	861	1 TOPI_BUCAI	P37371 buchnera ap
32	74.5	2.0	1021	1 CONT_RAT	Q63198 rattus norv
33	74.5	2.0	2203	1 CCAD_RAT	P27732 rattus norv
34	74	1.9	504	1 CAIT_ECOL6	P59333 escherichia
35	74	1.9	504	1 CAIT_ECOLI	P31553 escherichia
36	73.5	1.9	242	1 CRP3_LIMPO	P06207 limulus pol
37	73.5	1.9	362	1 RDC1_RAT	O89039 rattus norv
38	73.5	1.9	1018	1 CONT_HUMAN	Q12860 homo sapien
39	73	1.9	755	1 DNLI_YEAST	P04819 saccharomyc
40	73	1.9	776	1 CME3_BACSU	P39695 bacillus su
41	73	1.9	1070	1 YHV4_YEAST	P38850 saccharomyc
42	73	1.9	3565	1 CSM1_HUMAN	Q96p27 homo sapien
43	72.5	1.9	551	1 COX1_SYNP3	Q06473 synchocyst
44	72.5	1.9	1237	1 DP3A_CHLMU	Q9pj77 chlamydia m
45	72.5	1.9	1433	1 VGLM_BUNYW	P04505 bunyamwera

ALIGNMENTS

RESULT 1

ID	EP3A_HUMAN	STANDARD;	PRT;	147 AA.
AC	Q14507;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Epididymal secretory protein E3 alpha precursor (HE3 alpha).			
GN	FAM12A OR HE3ALPHA.			
OS	Homo sapiens (HUMAN).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Epididymis;			
RC	MEDLINE=94235297; PubMed=7514008;			
RA	Kirchhoff C., Pera I., Rust W., Ivell R.;			
RT	"Major human epididymis-specific gene product, HE3, is the first			
RL	representative of a novel gene family.";			
RL	Mol. Reprod. Dev. 37:130-137(1994).			
RN	[2]			
RP	REVISIONS.			
RA	Kirchhoff C.;			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Possible function in sperm maturation.			
CC	-!- SUBCELLULAR LOCATION: Secreted (Potential).			
CC	-!- TISSUE SPECIFICITY: Epididymis, with predominant expression in the			
CC	corpus region. Moderately expressed in the vas deferens; only low			
CC	levels are detectable in the caput and cauda regions.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X76383; CAA53971.2; -;
EMBL; X76385; -; NOT ANNOTATED_CDS.
Genew; HGNC:16978; FAM12A.

DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0007321; P:spERM displacement; TAS.
DR InterPro; IPR001427; RNaseA.
KW SIGNAL; Multigene family.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 147 EPIDIDYMAL SECRETORY PROTEIN E3 ALPHA.
SQ SEQUENCE 147 AA; 17646 MW; 3C700AEB01793A44 CRC64;

Alignment Scores:
Pred. No.: 0.604 Length: 147
Score: 87.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.27% Indels: 33
DB: 1 Gaps: 5

US-09-931-733-1 (1-2191) x EP3A_HUMAN (1-147)
QY 230 TCATCTATCAAAATTCATCTTTGGTATGAATATGAGCATGTGCTGAGT----- 283
Db 3 SerSerLeuLysile-----TrpGlyileuLeuAlaLeuLeu 15
QY 284 -----TGCGCGCTATAGTGACCGCAGTTTCTGTGTATGCT----- 319
Db 16 CysileLeuCysArgLeu-----CysValTyrSerAsnAsnileTyrTrp 30
QY 320 -----AGTCCCAATTGAGCATTTACTGTCATATAAATG 352
Db 31 ArgGluPheileLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
QY 353 ATTATAAGTAAAGGAGAGAAATGTCGAGGGTCATATTCGAGGGTCTCTAC 412
Db 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisMetPhe-Ile-- 69
QY 413 CTTTGACCTCTGGATTGGGGTTCCAGTGGAGTGCCTTAAT 452
Db 70 -TyrSerLeuTyrPheLysileGlnArgAlaCysileAsn 82

RESULT 2
HOLA_BUCAI STANDARD; PRT; 331 AA.
AC P57520;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase III, delta subunit (EC 2.7.7.7).
GN HOLA OR BU445.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT 'Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.';
RL Nature 407:81-86(2000).
CC -1- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The delta subunit seems to interact with the gamma subunit to
CC transfer the beta subunit on the DNA (By similarity).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -1- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC epsilon and theta chains) that associates with a tau subunit. This
CC core dimerizes to form the POLIII' complex. POLIII' associates
CC with the gamma complex (composed of gamma, delta, delta', psi and
CC chi chains) and with the beta chain to form the complete DNA
CC polymerase III complex (By similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

DR EMBL; AP001119; BAB13143.1; -
DR InterPro; IPR005790; DNA_PolIII_delta.
DR InterPro; IPR008921; Pol_clamp_load_C.
DR TIGRFAMs; TIGR01128; hola; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW Complete proteome.
SQ SEQUENCE 331 AA; 40359 MW; 4C87B19DA31493FD CRC64;

Alignment Scores:
Pred. No.: 0.724 Length: 331
Score: 87.00 Matches: 48
Percent Similarity: 40.18% Conservative: 40
Best Local Similarity: 21.92% Mismatches: 77
Query Match: 2.28% Indels: 54
DB: 1 Gaps: 12

US-09-931-733-1 (1-2191) x HOLA_BUCAI (1-331)
QY 621 AACCATATGAATACTTTATCAAAAAGGAGTAAGTCAAGAAAAGCAAGAGGATCTGG 562
Db 115 AsnHisLeuSerArgPheileGlnLysAsnLysSerLeuLysGluPheLysAsn----- 132
QY 561 AATATGGCCATAAATCATCTCTGGGTTCGCGTATTAATCTTTTACTGGTTCAGGTATCTC 502
Db 133 -----TyrAsnileValSerCys-----PheThrProTyrAsnLeuAsn 145
QY 501 TTT---ACGTGGATACATTGAGTTGGCCTTAAGCGAGTAGGTGAAGAGGATATTAAGGC 445
Db 146 PheileAsnTrpIle-----
QY 444 ACTCACTGGGACCCCAATCCAGAGGTCAAAGGGTAGAGAACCCCTTGAAGATTATACCC 385
Db 151 ---LysTyrGluileGlnGluLysLysileAsnileGluGlu----- 163
QY 384 TTCAGACATATTTCTGCTCTTTTACTT-----ATAATCAATTTATATCA 340
Db 164 ---LysAlaPhePheLeuLeuCysLysTyrTyrGluGlyAsnThrLeuPheileTyrLys 182
QY 339 GTAATGCTCAATTTGGGACTAGCATACACAAAATCGGTCCCTATAGCGGCAATCTC 280
Db 183 ileLeuAspMetLeuPheileileTyrProAspThrCysileThrGluLysLysileLys 202
QY 279 AGGACAATGCTCCAT-----AATTCATACCACAAAAGATGAATTTTCATA 235
Db 203 LysileileileGluPhePheAspValSerProSerTyrTrpIleAsnSerilePheGln 222
QY 234 GATGAAGCAAGAAATGCTTAAATAATGTC---ATTTCCTCCCTCTTTGTGGCAGTGCC 178
Db 223 GlyLysThrGluLysSerPheTyrileLeuAsnilePhe-----PheLysLysLysTyr 240
QY 177 AACATTTGAATTTGTGCTATCATACAGAAAAGTCAATAGTAGAATATACAAAATTCCT 118
Db 241 AsnProLeuileLeuValArgSerLeuGlnLysAspLeuLeuGlnLeuileHisMetLys 260
QY 117 GATAAGCAAAAATGTCGAATATATATATATATATATATATATATATATATATATATAC 58
Db 261 ArgGluLysLys-----IleSerileTyrValMet-----LeuGluLysTyr 274
QY 57 AGTGACTAAATTTATATTAATCAGGCATTTTCTGAATAAATCTCTCAATATATGTA 1
Db 275 Asn-----IlePheValThrArgLysPhePheileLysAlaPheAsnLysile 291

RESULT 3
BRC2_RAT

```

ID  BRCA2_RAT      STANDARD;      PRT; 3343 AA.
AC  O35923;
DT  15-MAR-2004 (Rel. 43, Created)
DT  15-MAR-2004 (Rel. 43, Last sequence update)
DT  15-MAR-2004 (Rel. 43, Last annotation update)
DE  Breast cancer type 2 susceptibility protein homolog.
GN  BRCA2.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Sprague-Dawley; TISSUE=Testis;
RX  MEDLINE=97384941; PubMed=9242436;
RA  McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA  Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;
RT  "Characterization of the rat and mouse homologues of the BRCA2 breast
RT  cancer susceptibility gene.";
RL  Cancer Res. 57:3121-3125(1997).
CC  -!- FUNCTION: May participate in a pathway associated with the
CC  activation of double-strand break repair and/or homologous
CC  recombination (by similarity).
CC  -!- SUBUNIT: Interacts with RAD51 (By similarity).
CC  -!- TISSUE SPECIFICITY: Highest expression in testis. Also expressed
CC  in spleen, skeletal muscle, thymus, mammary gland, heart, ovary,
CC  prostate, liver, lung, kidney and brain.
CC  -!- SIMILARITY: Contains 7 BRCA2 repeats.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; U89653; AAB71378.1; -.
DR  PIR; T42207; T42207.
DR  InterPro; IPR02093; BRCA2 repeat.
DR  InterPro; IPR008994; Nucleic_acid_OB.
DR  Pfam; PF00634; BRCA2; 8.
DR  PROSITE; PS50138; BRCA2_REPEAT; 6.
KW  Repeat.
FT  REPEAT 984 1018 BRCA2 1.
FT  REPEAT 1197 1231 BRCA2 2.
FT  REPEAT 1405 1439 BRCA2 3.
FT  REPEAT 1503 1537 BRCA2 4.
FT  REPEAT 1645 1669 BRCA2 5.
FT  REPEAT 1828 1845 BRCA2 6.
FT  REPEAT 1939 1973 BRCA2 7.
FT  REPEAT 2019 2053 BRCA2 8.
SQ  SEQUENCE 3343 AA; 372213 MW; 653DB110D2302A8D CRC64;

Alignment Scores:
Pred. No.: 1.37 Length: 3343
Score: 86.50 Matches: 49
Percent Similarity: 38.34% Conservative: 25
Best Local Similarity: 25.39% Mismatches: 57
Query Match: 2.27% Indels: 62
DB: 1 Gaps: 9

US-09-931-733-1 (1-2191) x BRCA2_RAT (1-3343)
QY 1482 AATGATATATACCAATAAGTTTTTCTTACAAATTCGATGCTGAAGAGGTTCC 1423
D 1315 AsnAlaTyrLys-----LeuGluAsnSerAspValSerLysSerSer 1328
QY 1422 AACATTAGAACTTTCTTACCAGCAGACTTACATGAGTGCATTAGAGACTTATATCCC 1363
D 1329 ThrSerGlyThr-----ValTyrIleAsn 1336
QY 1362 TTAATGATGCTCAATA-----TAAGAGAAGCAACATGGCTTTCTTCCCTCCT 1312

```

```

Db 1337 LysGlyAspSerAspLeuProPheAlaAlaGluLysGlyAsnLysTyr-----Pro 1353
QY 1311 TTCCTGTGCACAGAGTGGCAGTGGCAGGAGAGATCTTGTGTAGGAGCATCTATTGCTACC 1252
D 1354 GluSerCysThrGlnTyrValArgGluGluAsnAlaGlnIleLysGluSerValSerAsp 1373
QY 1251 AATATTCTTACGATCTCCCTGGTGGGAAATGCT--TCTATGAATCGCTATTGTT 1195
D 1374 LeuThrCysLeuGluValMetLysAlaGluGluThrCysHisMetLysSer----- 1390
QY 1194 TAAATACTACCTGAAGTATTTCATCAATATTATAACTGACAGCCTGGGCTTTCAGAG 1135
D 1391 -----SerAspLysGluGlnLeu-ProSe 1398
QY 1134 TCAGAAAGTCATTATGATATGTAATCTCTTAAATTTATATGAGAGGTTTACAC 1075
D 1398 rAspLysMetGluGlnAsn-----MetLysGluPheAsnI 1410
QY 1074 TAGCTGTCTCTTGACCATGACTAGTGTACTTTACCTCAATGATTAGATTCAAGATGCT 1015
D 1410 eSerPhe-----GlnThrAlaSerGlyLysAsnI 1420
QY 1014 CAGATTTTCTGCATTATCTCTGGATAAACAACGCTGGATAGTGCATCTCAACATGACAG 955
D 1420 eArgValSerLysGluSerLeuAsnLysSerValAsnIleLeuAspGlnGlu----- 1437
QY 954 CATCGAGAGCTTGAATCTCTGGAATGCACTGAAC 918
D 1438 -ThrGluAspLeuThrValThrSerAspSerLeuAsn 1449

RESULT 4
ID_OLFO_RAT STANDARD; PRT; 310 AA.
AC P23269;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Olfactory receptor-like protein I3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91191556; PubMed=1840504;
RA Buck L., Axel R.;
RT "A novel multigene family may encode odorant receptors: a molecular
RT basis for odor recognition.";
RL Cell 65:175-187(1991).
CC -!- FUNCTION: Putative odorant receptor.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Olfactory epithelium.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64385; AAA41748.1; -.
DR PIR; E23701; E23701.
DR InterPro; IPR000276; GPCR_Rhodpen.
DR Pfam; PF00001; 7tm1; 1.
DR PRINTS; PR00237; GPCRHOOPS.
DR PROSITE; PS00237; G PROTEIN RECEPT FL 1; 1.
DR PROSITE; PS0262; G PROTEIN RECEPT FL 2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 23 EXTRACELLULAR (POTENTIAL).

```



```

FT TRANSMEM 24 48 1 (POTENTIAL).
FT DOMAIN 49 55 2 (POTENTIAL).
FT TRANSMEM 56 77 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 78 98 3 (POTENTIAL).
FT TRANSMEM 99 118 4 (POTENTIAL).
FT DOMAIN 119 137 5 (POTENTIAL).
FT TRANSMEM 138 156 6 (POTENTIAL).
FT DOMAIN 157 194 7 (POTENTIAL).
FT TRANSMEM 195 217 8 (POTENTIAL).
FT DOMAIN 218 234 9 (POTENTIAL).
FT TRANSMEM 235 258 10 (POTENTIAL).
FT DOMAIN 259 270 11 (POTENTIAL).
FT TRANSMEM 271 290 12 (POTENTIAL).
FT DOMAIN 291 310 13 (POTENTIAL).
FT CARBOHYD 3 3 CARBOHYD (GLCNAC. . .) (POTENTIAL).
FT DISULFID 95 187 BY SIMILARITY.
SQ SEQUENCE 310 AA; 35247 MW; 9688233890205B19 CRC64;

Alignment Scores:
Pred. No.: 2.22 Length: 310
Score: 82.50 Matches: 56
Percent Similarity: 36.82% Conservative: 39
Best Local Similarity: 21.71% Mismatches: 62
Query Match: 2.16% Indels: 101
DB: 1 Gaps: 12

US-09-931-733-1 (1-2191) x OLF0_RAT (1-310)
QY 515 CCAGTAAAGATTATACGACACCCAGAGAGTGTATGGCCATTCACAGATCCTCTTT 574
Db 16 ProfileProGluHisGlnHis-----LeuPheTyraLeuPheLeuValMetTyr 33
QY 575 GCTTTTCTTGACATCTACTCTTTTTCATAGAGTATTCATATGTTTCAATACGAG--- 631
Db 34 -----LeuThrThrLeuGlyAsnLeuLeuIleValLeuValGlnLeu 49
QY 632 AATTCAGCATTCATCTACTGCGCTAGTTGGATTAAACAAACAGCTTCTTTTGTCTC 691
Db 50 AspSerGlnLeuHisThrProMetTyLeuPheLeuSerAsnLeuSerPheSerAspLeu 69
QY 692 TGT-----AGG 697
Db 70 CysPheSerSerValThrMetProLysLeuLeuGlnAsnMetArgSerGlnAspThrSer 89
QY 698 TGGCCTTGG-----ACATGATATTAACATCTATCGGCC----- 734
Db 90 IleProTyrglyGlyCysLeuAlaGlnThr-TyrPhePheMetValPheGlyAspMetG 109
QY 735 -----TTTCCACTCAT 745
Db 109 uSerPheLeuLeuValAlaMetAlaTyraArgTyraValAlaIleCysPheProLeuHi 129
QY 746 GACCACATCC-----TGCTTCGAGATCTGCTGCTGGGTAA 781
Db 129 sTyThrSerIleMetSerProLysLeuCysThrCysLeuValLeuLeuTyrMetLe 149
QY 782 CACCACATCATCTGCTAGT----- 800
Db 149 uThrThrSerHisAlaMetMetHisThrLeuLeuAlaAlaArgLeuSerPheCysGluAs 169
QY 801 -----TGATGCCATTATTTGACATCTCTCGGGGTTACAGCAGCTAA 841
Db 169 nAsnValValLeuAsnPhePheCysAspLeuPheValLeuLeuLysLeuAlaCysSerAs 189
QY 842 AGGCAC-----CCAGTGTCTTAAACAGAGAAAATACATGCTG 881
Db 189 pThrTyraAsnGluLeuMetIlePheIleMetSerThrLeuLeuIleIleIleProPh 209
QY 882 -----ACGACAGGATTAAACAGACTCTCTACACTACCGGAAGT 919
Db 209 ePheLeuLeuValMetSerTyraAlaArgIleSerSerIleLeuLysValProSerTh 229
QY 920 TCAGTGCATTTCCAGGATTTTCAAGTCT-----TCGATGCTGTCATGTT 964

```

```

Db 229 rGlnGlyIle-CysLysValPheSerThrCysGlySerHisLeuSerValValSerLeuP 249
QY 965 GAGAGTCCACTATC-----CACGTTGTTTATCCAGGAATAATGCA 1005
Db 249 heTyrglyThrIleIleGlyLeuTyLeuCysProAlaGlyAsnAsnSer 265

RESULT 5
NFIA_CHICK STANDARD; PRT; 522 AA.
ID NFIA_CHICK
AC P17923;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nuclear factor 1 A-type (Nuclear factor 1/A) (NFI-A) (NFI-1/A)
DE (CCAAT-box binding transcription factor) (CTF) (TGGCA-binding
DE protein).
DE NFIA OR NFI-A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=90251434; PubMed=2339052;
RA Rupp R.A.W., Kruse U., Multhaup G., Goebel U., Beyreuther K.,
RA Sippl A.E.;
RT "Chicken NFI/TGGCA proteins are encoded by at least three independent
RT genes: NFI-A, NFI-B and NFI-C with homologues in mammalian genomes.";
RL Nucleic Acids Res. 18:2607-2616(1990).
CC -/- FUNCTION: Recognizes and binds the palindromic sequence 5'-
CC TGGCANNNNCCAA-3', present in viral and cellular promoters and in
CC the origin of replication of adenovirus type 2. These proteins are
CC individually capable of activating transcription and replication.
CC -/- SUBCELLULAR LOCATION: Nuclear.
CC -/- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced;
CC Name=1;
CC IsoId=P17923-1; Sequence=Displayed;
CC -/- SIMILARITY: Belongs to the CTF/NFI family.
CC -/- SIMILARITY: Contains 1 DWA/MH1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51486; CAA35853.1; -.
CC PIR; S09996; S09996.
CC TRANSFAC; T00544; -.
CC InterPro; IPR000647; CTF NFI.
CC InterPro; IPR003619; Dwarfin A.
CC Pfam; PF00859; CTF NFI; 1.
CC Pfam; PF03165; MH1; 1.
CC SMART; SMO0523; DWA; 1.
CC PROSITE; PS00349; CTF NFI; 1.
CC Transcription regulation; DNA replication; DNA-binding; Activator;
KW Nuclear protein; Multigene family; Alternative splicing.
FT DOMAIN 67 175 DWA.
SQ SEQUENCE 522 AA; 57473 MW; C51F3E5FE6077740 CRC64;

Alignment Scores:
Pred. No.: 3.22 Length: 522
Score: 81.50 Matches: 60
Percent Similarity: 35.71% Conservative: 35
Best Local Similarity: 22.56% Mismatches: 92

```

```

Query Match: 2.13% Indels: 79
DB: 1 Gaps: 13
US-09-931-733-1 (1-2191) x NFIA_CHICK (1-522)
QY 412 CTTTGGACCTCTGGATTGGGG-----TTCCAGCTGAGTGCTTAAATATCCTTCTTCACC 465
DB 235 ProilealaalaGlyThrGlyProAsnPheSerLeuSerAspLeuGluSerSerTyr 254
QY 466 TACTGCCTTAAGGCAACTCAATGTATCCAGTAAAGAGTACCTTAAGCCAGT-----519
DB 255 TyrSerMetSerProGlyAlaMet-----ArgArgSerLeuProSerThrSerThr 272
QY 520 -----AAAGATTATCAGCAACACCAGACAGATGTTATGCCATATTCAGATCCCT 570
DB 273 SerSerThrLysArgGlyLeuSerValGluAspGluMetAspSerProGlyGluGluPro 292
QY 571 CTTTGGCTTTTCTTGACTTAC-----TCCTTT 597
DB 293 -----PheTyrThrSerGlnGlyArgSerProGlySerGlySerGlnSerSerGlyTyr 310
QY 598 TTTGATAAAGTATTATCATGGTTTCAATCCAGAAATTCAGTATTCATCTATGGCCCT 657
DB 311 HisGluValGluProGlyTyrLeuArgAsnProGluHisArgGlyAlaLeuHisGly---329
QY 658 AGTTTGGATTAAACAATCAGCTTCTTCTGCTAGATGGCTTGGACAGTACATAT 717
DB 330 -----MetProSerProThrAlaLeu 336
QY 718 TAACAGTCTAAGGCGCCCTTTCCACTCATGACCATCTGCTGTCAGATCTGCTGGG 777
DB 337 LysLysSerGluLysSerGlyPheSerProSerProSerGlnThrSerSerLeuGly 356
QY 778 -----TTAACACCATCATCTCTAGTTCATCCATTTATTTGACATCTCGGGTAC 831
DB 357 ThrAlaPheThrGlnHisArgProVal-----366
QY 832 AGCAGCTAAGGCGCCAGTCTTAAACAGAGAAATATACACTGCTGAGCAGAGAT 891
DB 367 -----IleThrGly-Pro-----370
QY 892 TAAACAGCTCTCACACTACCGGAAGTTCAGTGCATTTTCCAGGATTTTCAAGTCTTCG 951
DB 371 -ArgAlaSerProHisAlaThrProSerThrLeuHisPhePro---ThrSerProIle--388
QY 952 ATGCTGTCATGTTGAGAGTCCACTATCCAGTTGTTGTTATCCAGGAATA-----1000
DB 389 -----IleGlnGlnProGlyProTyrPheSerHisProAlaIleArgTyrHisPr 405
QY 1001 -ATGCAAAATCTGAGCATTCGTAATCTAATCATTCAGTAAAGTAACTAGTCATGG 1059
DB 405 oGlnGluThrLeuLysGluPheValGlnLeuValCys-----ProAspAlaG 421
QY 1060 TCAAGAGCAGCAGCTAGTGTAAATCTCTCATATAAATTTAAGGAAGGATTTACATATTC 1119
DB 421 yGlnGln-AlaGlyGlnValGlyPheLeuAsnProAsnGlySerSerGlnGlyLysValH 441
QY 1120 ATAATGACTTCTG 1133
DB 441 isAsnProPheLeu 445
RESULT 6
ID OLF8_RAT STANDARD; PRT; 312 AA.
AC P2327L;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Olfactory receptor-like protein I8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;

```

[1]
SEQUENCE FROM N.A.
MEDLINE=91191556; PubMed=1840504;
RA Buck L., Axel R.;
RT "A novel multigene family may encode odorant receptors: a molecular
basis for odor recognition.";
RL Cell 65:175-187(1991).
CC -1- FUNCTION: Putative odorant receptor.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Olfactory epithelium.
CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; M64387; AAA41750.1; --
DR PIR; G23701; G23701.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHDOPS.
DR PROSITE; PS00237; G PROTEIN RECP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECP_F1_2; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 23 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 24 48 1 (POTENTIAL).
FT DOMAIN 49 55 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 56 77 2 (POTENTIAL).
FT DOMAIN 78 98 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 99 118 3 (POTENTIAL).
FT DOMAIN 119 137 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 138 156 4 (POTENTIAL).
FT DOMAIN 157 194 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 195 217 5 (POTENTIAL).
FT DOMAIN 218 234 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 235 258 6 (POTENTIAL).
FT DOMAIN 259 270 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 271 290 7 (POTENTIAL).
FT DOMAIN 291 312 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 95 187 BY SIMILARITY.
SQ SEQUENCE 312 AA; 35315 MW; D84D101F2DB6C29A CRC64;
Alignment Scores:
Pred. No.: 4.19 Length: 312
Score: 80.00 Matches: 39
Percent Similarity: 34.48% Conservative: 21
Best Local Similarity: 22.41% Mismatches: 39
Query Match: 2.09% Indels: 75
DB: 1 Gaps: 7
US-09-931-733-1 (1-2191) x OLF8_RAT (1-312)
QY 533 GAACACCCAGAGATGTTATGGCCATATTCAGATCCTCTTTGCTTTTCTTCTGACTTACT 592
DB 20 GluHisGlnGlnLeuPhePheAlaLeuPheLeuMetTyr-----LeuThr 35
QY 593 CCTTTTGTGATAAAGTATTCATATGTTTCAATCCAG-----AATTCAGTATTCTCTTC 649
DB 36 ThrPheLeuGlyAsnLeuLeuValValLeuValGlnLeuAspSerHisLeuHisThr 55
QY 650 ATGGCCCTAGTTGGATTAAACAATCAGCTTTCTTTTCTCTCTGT-----694
DB 56 ProMetTyrLeuPheLeuSerAsnLeuSerPheSerAspLeuCysPheSerSerValThr 75
QY 694 -----694
DB 76 MetLeuLysLeuLeuGlnAsnIleGlnSerGlnValProSerIleSerTyrAlaGlyCys 95

QY 695 -----AGATGGCCCTGGACAGTACATATTAAACAGT 724
 Db 96 LeuThrGlnIlePhePheLeuLeuPheGlyTyr-LeuGlyAanPheLeuLeuValAl 115
 QY 725 CTATGGG-----CCCTTCCACTCATGACCACA----- 752
 Db 115 aMetAlaTyrAspArgTyrValAlaIleCysPheProLeuHisTyrThrAsnIleMetSe 135
 QY 753 -----TCCTGCTTCAGATCGTCTGGGTAAACACACATCATCTGCTAG 799
 Db 135 rHisLysLeuCysThrCysLeuLeuValPheTrpIleMetThrSerSer----- 152
 QY 800 TTGATGCCATTATTGATCATCTCGGGTTACAGCAGCTAAAGGCGCCCA----- 851
 Db 153 -----HisAlaMetMetHisThrLeuLeuAl 161
 QY 852 -----GTGCTCTTAACAGAGAAAATACATGCTGAGC 884
 Db 161 aAlaArgIleSerPheCysGluAsnValLeuLeuAsn 174
 RESULT 7
 ID GP18 HUMAN STANDARD; PRT; 331 AA.
 AC Q14330; Q96H16;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable G protein-coupled receptor GPR18.
 GN GPR18 OR GPCR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97349114; PubMed=9205118;
 RA Gantz H., Murakami A., Yang Y.-K., Samuelson L.C., Zimmerman E.M.,
 RA Cook H., Yamada T.;
 RA "Cloning and chromosomal localization of a gene (GPR18) encoding a
 RT novel seven transmembrane receptor highly expressed in spleen and
 RT testis";
 RL Genomics 42:462-466 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOKINE RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN TESTIS AND SPLEEN.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L43224; AAB65819.1; --
 DR EMBL; BC008569; AAB08569.1; --
 DR Genbank; HGNC:4472; GPR18.
 DR MIN; 602042; --
 DR GO; GO:0016021; C: integral to membrane; NAS.
 DR GO; GO:0004930; F: G-protein coupled receptor activity; NAS.
 DR GO; GO:0007186; P: G-protein coupled receptor protein signaling. .; NAS.
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PR00237; GPCRHOPOPSN.
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR F1_1; 1.
 DR PROSITE; PS0262; G-PROTEIN RECEPTOR F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 26
 FT TRANSMEM 27 47
 FT DOMAIN 48 56
 FT TRANSMEM 57 77
 FT DOMAIN 78 95
 FT TRANSMEM 96 116
 FT DOMAIN 117 138
 FT TRANSMEM 139 159
 FT DOMAIN 160 191
 FT TRANSMEM 192 212
 FT DOMAIN 213 232
 FT TRANSMEM 233 253
 FT DOMAIN 254 268
 FT TRANSMEM 269 289
 FT DOMAIN 290 331
 FT CARBOHYD 14 14
 FT CONFLICT 12 12
 FT CONFLICT 86 86
 FT CONFLICT 97 97
 FT CONFLICT 310 310
 FT CONFLICT 331 331
 SQ SEQUENCE 331 AA; 38061 MW; 95C9C33BE25D7606 CRC64;
 Alignment Scores:
 Pred. No.: 4.25 Length: 331
 Score: 80.00 Matches: 28
 Percent Similarity: 39.52% Conservative: 21
 Best Local Similarity: 22.58% Mismatches: 47
 Query Match: 2.10% Indels: 28
 DB: 1 Gaps: 5
 US-09-931-733-1 (1-2191) x GP18_HUMAN (1-331)
 QY 603 ATCAAAAAGAGTAAAGTCAAGAAAACAAAGAGGATCTGGATATGCGCATAAACATC 544
 Db 222 LeuLysProLysValLysGluLysSerLeuArgIleIleThrLeuLeuValGlnVal 241
 QY 543 TCTGGGTGTCTGATATCTTTTACTGGTCTTACTGGTCTTACTGGTCTGATACATGA 484
 Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255
 QY 483 GTTGGCCCTTAAGGCGAGTAGGTGAAGAGGATATAAGGCACCTCACTGGGAACCCCAATCC 424
 Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264
 QY 423 AGAGTCAAGAGGTAGAGAACCTTGGAGGTATGACCCCTTCAGAACATTTTCT3CCTT 364
 Db 265 -----AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn 276
 QY 363 TTA-----CTTATAATCATTTTATATGTCAGTAAATGTCTCAATTCGGAGTAGCATAC 313
 Db 277 LeuSerThrCysLeuAspValIleLeuTyrThrIleValSerLysGlnPheGlnAlaArg 296
 QY 312 ACACAACTGGCGTCCACCTATAGGGCGCACTACAGGACATGCTCCATTAATTCATACAC 253

```

Db      217  LysSerAspGylIleGlnPro--IleGlyLeuGlyAlaArgAspThrLeuArgLeuGlu 235
QY      512  TTAGGGTACTCTTTACCGTGNATAC-----ATTGAGTTGGCC 477
Db      236  MetGlyTyrSerLeuTyrGlyHisGluLeuArgAspValAsnProLeuGluAlaArg 255
QY      476  TTAAGGCAGTAGGTGAAGAAGATATTAAAGCCACTCACTGGG----- 435
Db      256  LeuLysTrpValValLysLeuAsnLysProAsnPheIleGlyLysGlnAlaCysGluGln 275
QY      434  -----AACCCCAATCCAGAGGTCAAAGGTTAGAGAACCTTGGAGTTATGACCCCT 384
Db      276  ValGluIleAsnProArgLysSerValValGlyPheSerLeuGluGlyArgAlaIlePro 295
QY      383  TCAGAACACATTTTCTGCCTTT 363
Db      296  ArgGlnHisPheLysValTyr 302

RESULT 9
GRAB_RAT      STANDARD;      PRT;      258 AA.
ID   GRAB_RAT
AC   P49864;
DT   01-OCT-1996 (Rel. 34, Created)
DI   01-OCT-1996 (Rel. 34, Last sequence update)
DE   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Granzyme K precursor (BC 3.4.21.-) (NK-tryptase-2) (NK-TRYP-2).
GN   GZMK
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 26-58.
RC   STRAIN=Fischer 344; TISSUE=Lymphocytes;
RX   MEDLINE=94179809; PubMed=8133042;
RA   Sayers T.J., Wiltrout T.A., Smyth M.J., Ottaway K.S., Pilaro A.M.,
RA   Sowder R., Henderson L.E., Sprenger H., Lloyd A.R.;
RT   "Purification and cloning of a novel serine protease, RNK-Tryp-2,
RT   from the granules of a rat NK cell leukemia.";
RT   J. Immunol. 152:2289-2297(1994).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic granules.
CC   -!- TISSUE SPECIFICITY: Spleen, lungs and liver nonparenchymal cells.
CC   -!- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
Db      EMBL; L19694; AAA42057.1; -.
Db      PIR; I56220; I56220.
Db      HSP; P20160; 1A7S.
Db      MEROPS; S01.146; -.
Db      InterPro; IPR009003; Cys_Ser_trypsin.
Db      InterPro; IPR001254; Peptidase_S1.
Db      InterPro; IPR001314; Peptidase_S1A.
Db      Pfam; PF00089; trypsin; 1.
Db      PRINTS; PR00722; CHYMOTRYPSIN.
Db      SMART; SM00020; TRYD_SPC; 1.
Db      PROSITE; PS0240; TRYPSIN_DOM; 1.
Db      PROSITE; PS00134; TRYPSIN_HIS; 1.
Db      PROSITE; PS00135; TRYPSIN_SER; 1.
Db      KXW Hydrolase; Serine protease; Zymogen; Signal.
Db      SIGNAL 1 23 POTENTIAL.
Db      PROPEP 24 25 ACTIVATION PEPTIDE.
Db      FT CHAIN 26 258 GRANZYME K.
Db      FT ACT_SITE 56 66 CHARGE RELAY SYSTEM (BY SIMILARITY).
Db      FT ACT_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).
Db      FT ACT_SITE 208 208 CHARGE RELAY SYSTEM (BY SIMILARITY).

```



```

Db 232 LysGluLeuValSerPheThrGlnLysThrTyrSerAspLeu-----IleLeuArgPheProGlu 250
QY 1869 TGGAACTTTACTGATTTATCTCCCTCCCTCACAAATAACATGATCTCTGCAATTTCT 1928
Db 251 GluGluLeuLeuMetLeuValSerSerPheGlnSerLeuIleTyrAsnGluPheValSer 270
QY 1929 GAA-----GATCTCAAGATCTGGATCTGTTGAAAAA-----1962
Db 271 GluIlePheIleSerAsp-AnpPheThrGlyValTyrPheLysThrLysThrGlyProLe 290
QY 1963 -----TTTCCAGGTGAG 1974
Db 290 uPhePheProGlyGlu 295
RESULT 11
SHT_HELVI
ID SHT_HELVI STANDARD; PRT; 466 AA.
AC Q25190;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 5-hydroxytryptamine receptor (5-HT receptor) (Serotonin receptor).
OS Heliothis virescens (Noctuid moth) (Owllet moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Heliothinae; Heliothis.
OX NCBI_TaxID=7102;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Antenna;
RX MEDLINE=97166608; PubMed=9014328;
RA von Nickisch-Rosenegk E., Krieger J., Kubick S., Laage R., Strobel J.,
RA Strotmann J., Breer H.;
RT "Cloning of biogenic amine receptors from moths (Bombyx mori and
RT Heliothis virescens).";
RL Insect Biochem. Mol. Biol. 26:817-827(1996).
CC -1- FUNCTION: This is a receptor for 5-hydroxytryptamine (serotonin),
CC a biogenic hormone that function as a neurotransmitter, a hormone,
CC and a mitogen (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X95605; CAA64863.1;
CC InterPro; IPR000276; GPCR_Rhodpsn.
CC Pfam; PF00001; 7tm1_1;
CC PRINTS; PR00237; GPCRHHODPSN.
CC PROSITE; PS00237; G PROTEIN RECEPT FL 1; 1.
CC PROSITE; PS0262; G PROTEIN RECEPT FL 2; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 66 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 67 89 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 90 99 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 100 121 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 122 136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 137 158 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 177 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 178 200 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 201 228 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 229 250 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 251 386 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 387 410 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 411 419 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 420 442 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 443 466 CYTOPLASMIC (POTENTIAL).

```

```

FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 29 29 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 50 50 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 135 215 BY SIMILARITY.
SQ SEQUENCE 466 AA; 52094 MW; AD29A38EC37106EB CRC64;

Alignment Scores:
Pred. No.: 5.9 Length: 466
Score: 79.00 Matches: 34
Percent Similarity: 40.31% Conservative: 18
Best Local Similarity: 26.36% Mismatches: 61
Query Match: 2.07% Indels: 16
DB: 1 Gaps: 6

US-09-931-733-1 (1-2191) x SHT_HELVI (1-466)
QY 1605 TGGGGGATGAGAAATTAATTAATGGGTACAGTAAATTTTCGGGTGATGTGCTCACT 1546
Db 38 TrpGlyCysAsnValThrIleAsnGlyThrSer-----ThrAsnAlaThrSerThr 54
QY 1545 GAAATCTCTGACTTACCAATATGGCATATATCCATGTAAACGAAATTTACACTTGTATCTC 1486
Db 55 AspValThrSerPheValLeuMetAlaValThrSerValValLeuAlaLeuIleLeu 74
QY 1485 ATAATGTATATAACACAAATAAGTTTTTTTCTTACAAATTCGATAGCTGAAGAGT 1426
Db 75 AlaThrIleValGlyAsn-----ValPheValIleAlaIleIleGluArgAsn 92
QY 1425 ---TCCAACTTAGAATTTCTTACCAGCAGACTTACATGTGCTAGCTAGTAGACT--- 1372
Db 93 LeuGlnAsnValAlaAsnFyrLeuValAlaSerLeuAlaValAlaAspLeuMetValAla 112
QY 1371 ----TATATCCCTTAAATGATGCTCAATATAAGAGAGCAACATGCTCTTCTCTCC 1318
Db 113 CysLeuValMetProLeu-----GlyAlaValTyrGluValSerGlnGlyTyrPheLeu 130
QY 1317 CCTCTTCTGTCACAGAGTGGCAGAGGAGGAGGATCTTGGTAGGAGCATCTATT 1258
Db 131 GlyProGluLeuCysAspMetTyrThrSerSerAspValLeuCysSerSerAlaSerIle 150
QY 1257 GCTACCAATACTTACTGCTCTCTG 1231
Db 151 -----LeuHisLeuVal 154

RESULT 12
PRP5_SCHPO
ID PRP5_SCHPO STANDARD; PRT; 473 AA.
AC Q13615;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Pre-mRNA splicing protein prp5.
GN PRP5 OR CWF1 OR SPBP22H7.07 OR PI024.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 26-47, AND FUNCTION.
RX MEDLINE=99340235; PubMed=10409726;
RA McDonald W.H., Ohi R., Smelkova N., Frendewey D., Gould K.L.;
RT "Myb-related fission yeast cdc5p is a component of a 40S
RL snRNP-containing complex and is essential for pre-mRNA splicing."
RN Mol. Cell. Biol. 19:5352-5362(1999).
RP [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=20089027; PubMed=10620777;
RA Machida M., Yamazaki S., Kunihiro S., Tanaka T., Kushida N., Jinno K.,

```


RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriuni M.J., Town C.D., Walker M.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
 RT *thaliana*.";
 RL Nature 408:816-820(2000).
 RN [3]
 RN FUNCTION, AND MUTANTS SELL.
 RX MEDLINE=21836422; PubMed=11846880;
 RA Shibagaki N., Rose A., McDermott J.P., Fujiwara T., Hayashi H.,
 RA Yoneyama T., Davies J.P.;
 RT "Selenate-resistant mutants of *Arabidopsis* thaliana identify Sultr1;2,
 RT a sulfate transporter required for efficient transport of sulfate into
 RT roots.";
 RL Plant J. 29:475-486(2002).
 CC -!- FUNCTION: High-affinity H(+) /sulfate cotransporter that mediates
 CC the uptake of the environmental sulfate by plant roots. Plays a
 CC central role in the regulation of sulfate assimilation.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- TISSUE SPECIFICITY: Expressed in lateral root cap, root hairs,
 CC epidermal and cortical cells of roots.
 CC -!- INDUCTION: In roots by sulfate starvation.
 CC -!- MISCELLANEOUS: Sell mutations in the gene lead to the resistance
 CC of the plant to selenate, a toxic analog of sulfate.
 CC -!- SIMILARITY: Belongs to the SLC26A/Sulph transporter (TC 2.A.53)
 CC family.
 CC -!- SIMILARITY: Contains 1 STAS domain.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AB042322; BAA95484.1;
 DR EMBL; AC009243; AAF17685.1; ALT_SEQ.
 DR PIR; D96809; D96809.
 DR InterPro; IPR002645; STAS.
 DR InterPro; IPR01902; Sulph_transp.
 DR Pfam; PF01740; STAS; 1.
 DR Pfam; PF00916; Sulfate_transp; 1.
 DR TIGRfam; TIGR00815; sulp; 1.
 DR PROSITE; PS01130; SLC26A; 1.
 DR PROSITE; PS00801; STAS; 1.
 DR Transport; Symport; Sulfate transport; Transmembrane;
 KW Multigene family.
 KW DOMAIN 1 91 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 92 112 POTENTIAL.
 FT DOMAIN 113 116 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 117 137 POTENTIAL.
 FT DOMAIN 138 141 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 142 162 POTENTIAL.
 FT DOMAIN 163 173 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 174 194 POTENTIAL.
 FT TRANSSEM 195 215 POTENTIAL.
 FT DOMAIN 216 253 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 254 274 POTENTIAL.
 FT DOMAIN 275 280 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 281 301 POTENTIAL.
 FT DOMAIN 302 339 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 340 360 POTENTIAL.
 FT DOMAIN 361 372 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 373 393 POTENTIAL.
 FT DOMAIN 394 409 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 410 430 POTENTIAL.
 FT DOMAIN 431 438 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 439 459 POTENTIAL.
 FT DOMAIN 460 466 EXTRACELLULAR (POTENTIAL).
 FT

FT TRANSSEM 467 487 POTENTIAL.
 FT DOMAIN 488 553 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 522 645 STAS.
 FT MUTAGEN 511 511 I->T: IN SELL-8; REDUCES DRAMATICALLY
 FT SULFATE TRANSPORT ACTIVITY.
 SQ SEQUENCE 653 AA; 71707 MW; B019CADD94DED46 CRC64;
 Alignment Scores:
 Pred. No.: 7.21 Length: 653
 Score: 78.50 Matches: 28
 Percent Similarity: 40.91% Conservative: 17
 Best Local Similarity: 25.45% Mismatches: 40
 Query Match: 2.06% Indels: 25
 DB: 1 Gaps: 4
 US-09-931-733-1 (1-2191) x STL2_ARATH (1-653)
 QY 1296 TGGCAGTGGCAGGAGAGATCTTGGTAGGAGCATCTATTGCTACCAATACCTTTACAT 1237
 Db 252 TrpAsnTrpGln---ThrIleLeuIleGlyAlaSerPheLeuLeuThrSer 270
 QY 1236 CTCCTGGTGGGGAATGTCT-----TCTATGAATCCGATTT 1198
 Db 271 LysIleIleGlyLysLysSerLysLeuPheTrpValProAlaIleAlaProLeuIle 290
 QY 1197 GTTTAATACTACTGAGTATTCATCCATTAATTAATACTGACACGCTGGGCTTCC 1138
 Db 291 SerValIleValSerThrPhePheValIleThrArgAlaAspLysGlnGlyValGln 310
 QY 1137 AGATCAGAAAGTCATTAATGATATGTAATCCCTCCCTTA----- 1099
 Db 311 IleValLysHisLeuAspGlnGlyIleAsnProSerSerPheHisLeuIleTyrPheThr 330
 QY 1098 -----AATTATATGAGGAGTTTAACTAGCTGTCTCTTGACCATGACTAGTTAC 1045
 Db 331 GlyAspAsnLeuAlaLysGlyIle-----ArgIleGlyVal 342
 QY 1044 TTTACCTCAATGATTAGATTCACGAAATGCT 1015
 Db 343 ValAlaGlyMetValAlaLeuThrGluAla 352
 RESULT 14
 CABB HUMAN
 ID CABB HUMAN STANDARD; PRT; 252 AA.
 AC P2085; 1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE C4b-binding protein beta chain precursor.
 GN C4BPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=93315479; PubMed=8325897;
 RA Hillarp A., Pardo-Manuel F., Ruiz R., de Cordoba S., Dahlback B.;
 RT "The human C4b-binding protein beta-chain gene.";
 RL J. Biol. Chem. 268:15017-15023(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90138962; PubMed=2300577;
 RA Hillarp A., Dahlback B.;
 RT "Cloning of cDNA coding for the beta chain of human complement
 RT component C4b-binding protein: sequence homology with the alpha
 RT chain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1183-1187(1990).
 CC -!- FUNCTION: C4BP controls the classical pathway of complement
 CC activation. It binds as a cofactor to C3b/C4b inactivator
 CC (C3bINA), which then hydrolyzes the complement fragment C4b. It
 CC also accelerates the degradation of the C4bC2a complex (C3
 CC convertase) by dissociating the complement fragment C2a. It also

Alignment Scores:

Pred. No.:	7.38	Length:	409
Score:	78.00	Matches:	36
Percent Similarity:	36.00%	Conservative:	18
Best Local Similarity:	24.00%	Mismatches:	40
Query Match:	2.05%	Indels:	56
DB:	1	Gaps:	6

US-09-931-733-1 (1-2191) x AMB1_BACST (1-409)

Qy	1509	GTACGAAATACACTTCTATCTC-----ATAAATGTATATAAACAATAAAGTTTTT	1456
Db	262	ValGlyGlnLeuHisValTyrProGlyGlyIleAsnValIleProGluArgValGluPhe	281
Qy	1455	TTCTTACAA-----AATTCGATAGCT	1435
Db	282	ValLeuAspLeuArgAspLeuLysAlaGluValArgAspGlnValTrpLysAlaIleAla	301
Qy	1434	GAAGAGGAGTCCCAACATTAGAACTTCTTACCCAGCAGACTTACATGTGAGTCATTAGAG	1375
Db	302	ValArgAlaGluThrIleAlaLysGluArgAsnValArgLeuThrGluArgLeuGln	321
Qy	1374	ACTTATATTCCTTAAATGATGCTCAAATATAAGAG-----AAG	1336
Db	322	GluMetAlaProValLeuCysSerGluValValLysGlnAlaGluArgAlaCysLys	341
Qy	1335	CAACATGGCTTCTTCTCCCTCCTTTC-----	1309
Db	342	GlnLeuGlyTyr-----ProProPheTrpLeuProSerGlyAlaAlaHisAspGlyVal	359
Qy	1308	-----CTGTGC-----	1303
Db	360	GlnLeuAlaProIleCysProIleGlyMetIlePheValArgSerGlnAspGlyValSer	379
Qy	1302	-----ACAGATGGCAGTGGCAGGAGGAGATCTTGTGAGGAGCATCTATTGCTACC	1252
Db	380	HisSerProAlaGluTrpSerThrLysGluAspCysAlaValGlyAlaGluValLeuTyr	399
Qy	1251	AATACTTTACTGCATCTCCTGTTGGGAA	1222
Db	400	HisThrValTrpGlnLeuAlaGlnGlyGlu	409

Search completed: March 28, 2004, 08:06:16

Job time : 44 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 07:42:51 ; Search time 164.5 Seconds
(without alignments)
8404.868 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cpn2.1/USPTO.spool/US09931733/runat_25032004.155917.9648/app_query.fasta.1.2375
-DB=SPREMBL_25 -QFWT=fascan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733 @CNC 1.1 187 @runat_25032004.155917.9648 -NCPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_25:*
1: sp archaea.*
2: sp bacteria.*
3: sp fungi.*
4: sp human.*
5: sp invertebrate.*
6: sp mammal.*
7: sp mhc.*
8: sp organelle.*
9: sp phase.*
10: sp plant.*
11: sp rodent.*
12: sp virus.*
13: sp vertebrate.*
14: sp unclassified.*
15: sp_rvirus.*
16: sp bacteriap.*
17: sp archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	2.7	3290	10 Q9SH73	Q9sh73 arabidopsis

2	98	2.6	175	16	Q8ZG02
3	98	2.6	220	16	Q8CKX8
4	94	2.5	3135	5	Q813B5
5	89.5	2.3	306	3	Q870P3
6	86.5	2.3	374	6	Q8H2R7
7	86.5	2.3	3343	11	Q35923
8	85	2.2	409	2	Q8GQ55
9	84.5	2.2	230	16	Q8PAM4
10	83.5	2.2	507	5	Q9GTJ5
11	83	2.2	333	4	Q9H2L2
12	83	2.2	852	5	Q81B40
13	83	2.2	1208	16	Q7VRE9
14	82.5	2.2	396	15	Q85483
15	82.5	2.2	446	8	Q8HD85
16	82.5	2.2	600	15	Q83129
17	82	2.2	648	16	Q82Z22
18	82	2.1	676	3	Q9P7E5
19	82	2.2	837	13	Q7T1T7
20	82	2.2	921	5	Q7YZ56
21	81.5	2.1	479	11	Q7TSF9
22	81.5	2.1	582	5	Q9N515
23	81	2.1	205	16	Q933A7
24	81	2.1	866	11	Q80XL8
25	80.5	2.1	389	16	Q83E71
26	80.5	2.1	443	8	Q34844
27	80	2.1	302	5	Q81471
28	80	2.1	478	10	Q9SAGA
29	80	2.1	740	10	Q93325
30	80	2.1	1419	5	Q813G5
31	79.5	2.1	272	4	Q95992
32	79.5	2.1	330	16	Q8E4H8
33	79.5	2.1	330	16	Q8DYW9
34	79	2.1	250	16	Q8YXL3
35	79	2.1	310	11	Q7TRX5
36	79	2.1	335	3	Q9P6N1
37	78.5	2.1	396	2	Q9EVD2
38	78.5	2.1	2954	12	Q96898
39	78	2.0	251	4	Q9BS25
40	78	2.0	292	16	Q814H9
41	78	2.0	312	11	Q8VGR6
42	78	2.0	314	16	Q889G9
43	78	2.0	333	8	Q9MGE0
44	78	2.0	335	12	Q91HQ2
45	78	2.0	409	16	Q7U8C3

ALIGNMENTS

RESULT 1
Q9SH73
ID Q9SH73 PRELIMINARY; PRT; 3290 AA.
AC Q9SH73
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE F22C12.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Khan S., Brooks S., Buehler E., Chao Q., Dunn P., Kim C.,
RA Walker M., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F22C12 from chromosome
I.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007764; AAF24584.1; -;
DR GO; GO:0005524; F:ATP binding; IEA.

RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.,
 RT "Sequence of Plasmidium falciparum chromosomes 1, 3-9 and 13."
 RL Nature 419:527-531(2002).
 DR EMBL; AL929355; CAD51718.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 3135 AA; 372458 MW; 40548E3A4123545F CRC64;

Alignment Scores:
 Pred. No.: 0.298 Length: 3135
 Score: 94.00 Matches: 62
 Percent Similarity: 36.08% Conservative: 43
 Best Local Similarity: 21.31% Mismatches: 101
 Query Match: 2.46% Indels: 85
 DB: 5 Gaps: 15

US-09-931-733-1 (1-2191) x Q813B5 (1-3135)

QY 1 TACATAATGAGAGATTTTATTCAGATAATGCTGATTAATAATAAATTTAGTCACTGTA 60
 Db 622 TyrPheValAspLys--TyrSerIleSerCysLeuIleTyrMetAsnIleTyrThrGlu 640
 QY 61 ATAATGAGCTACATAATATATA-----TATGTATAT-----ATA 96
 Db 641 GluSerIleIleIleLysTyrIleIleThrIleLeuPheTyrIleHisCysAspLeuPhe 660
 QY 97 TTCGCACATTTTTCGTTATCAGGAATTTTGTATATCTACTATGACTTTCTGTATG 156
 Db 661 PheLysHisIleSerTyrLeuGluAsnMet-----Tyr 671
 QY 157 ATAGCAAAATACAAATGTTGGCCACTGCCACAAAGTGGGGGAAATGACATTTTAA 216
 Db 672 IleSerLysThrAsnIleSerLeuTyrCysTyrLys-----GluSerHisHisIleSer 689
 QY 217 GCATTT-----CTTTGCTTCATCTATGAA 240
 Db 690 LeuTyrAsnTyrIleLysLysIleIleSerSerIlePheLeuPheCysPheIleTyrSer 709
 QY 241 AATTTTC---ATCTTT---GTGTATGAATATGAGACATTTCTCCTGAGTCCGCGCT 291
 Db 710 GluPheAspIleTyrGlnAsnValIleTyrPheProLysHisLeuHisAsnIle----- 727
 QY 292 ATAGGTGACCGCAGTTTGTGTATGCTAGTCCCAATGAGACATTTACTGCATATAAAAT 351
 Db 728 -----ThrHisIleLeuHisPheCys----- 734
 QY 352 GATTTAATGATAAAGGAGAAAATGTTCTGAAGGGTCATTAACCTCCAGGGTCTCTAC 411
 Db 735 AspGluProTyrIleSerSerSerCysIleGlnSer----- 747
 QY 412 CCTTTGACCTCGAATGGGGTTCACAGTGAGTCCCTTAATATCCTTCTTCCACTAC--- 468
 Db 748 -----IleTyrPhePheLeuTyrLys 754
 QY 469 -----TGCCTTAAGGCAACTCAATGATCCAGTAAGAGTACCCTAAAGCCAGTAAA 522
 Db 755 TyrIleCysTyrHisPhePheSerIlePheIleGlnAsnIleLeuTyrLysIleMetHis 774
 QY 523 AGATTATACGAACACCCAGAGATGTTTATGGCCATATTCAGATCCTCTT----- 573
 Db 775 SerThrSerProIleArgHisGluGluAsnAspHisIlePheTyrProLeuSerValLeu 794
 QY 574 -----TGCTTTTCTTGACTTACTCTTTTGTAAAGTTATTTCATATGTT 621
 Db 795 LysIleMetLysCysPheAsnTyrTyrTyrAspTyrPheSerAsnIlePheIleArgIle 814
 QY 622 TCAATCCAGAAATTCAGCTATTCACCTTCATGCGCCCTAGTTTGGATTAAACCAATCAGTT 681

Db 814 uAsnIleLysMetAsnLeuIleLeuLysGluArgCysLeuIleMetLysAspIleGluAs 834
 QY 682 TCTTTTCTCTCTAGATGGCTTGGACAGTACATATT-----AACAGTCTATGGGCCCT 735
 Db 834 nIleLysIleLeuIleMetProLeuThrTyrAsnIleValCysAsnAsnMetTyrAsnTy 854
 QY 736 TTCC-----ACTCATGACCATCTCGCTT 760
 Db 854 rGlyHisMetGluHisAspHisPheIleVal 864

RESULT 5
 Q870P3 PRELIMINARY; PRT; 906 AA.

AC Q870P3
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein 49D12.100.
 GN 49D12.100.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Hollanc R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX295540; CAD79685.1; -
 DR InterPro; IPR006598; CAP10.
 DR SMART; SM00672; CAP10; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 906 AA; 102137 MW; 71EDE49AF8E03146 CRC64;

Alignment Scores:
 Pred. No.: 0.786 Length: 906
 Score: 89.50 Matches: 60
 Percent Similarity: 39.18% Conservative: 36
 Best Local Similarity: 24.49% Mismatches: 84
 Query Match: 2.34% Indels: 66
 DB: 3 Gaps: 14

US-09-931-733-1 (1-2191) x Q870P3 (1-906)

QY 271 CATTTCTCTGAGTTGCGCCTATAGGTGACGCGAGTTTGTGTATGCTAGTCCCATTTG 330
 Db 550 HisCys---SerMetProProLysArgGluSerAsnProLeuSerCys----- 564
 QY 331 AGACATTACTGCATATAAAATGATTATAAGTAAAGCAGAAAAATTTCTG----- 382
 Db 565 LysHis-AspProPheLysLeuLeu-----LysArgGlnLeuThrThrLeuAlaIyrG 582
 QY 383 -----AAGGTCATACCTTCCAGGGTCTCTACCTTTGACCTCTGGA 426
 Db 582 uProGluCysLeuThrLysGlyValProPheGln----- 593
 QY 427 TTGGGGTTCACAGTGAGTGCCTTAATATCTTCTTCCACTACTGCTTAAAGGCCAACTCA 486
 Db 594 -----AsnSerProAlaIleValProSerG 603
 QY 487 ATGTATCCAGTAAAGAGTACCCTAAAGCCAGTAAAGATTTATCAGAACACCCAGAGAT 546
 Db 603 uSerArgHisAlaArgSer-----ProValLeuProTyrProProHis----- 617
 QY 547 GTTTATGGCCATATTCAGATCCTTTTGTCT-----TTTCTTGACTTACTCTTTT 600
 Db 618 -----IleLeuArgGluLeuPheAlaArgAspPheLeuSerAlaCysProArg 634
 QY 601 GATAAAGTTATTTCATATGTTT-----CAATCCAGAAATTCACCTAT 642

```

Db      634 sArgArgMetLeuValTrpIleLysLysGlyValLeuGlnGlnLeuAspValArgArgLe 654
QY      643 TCAC-----TTTCATGGCCCTAGTTTGGATTAAACATCAGCTTCTTTT 687
Db      654 uHisArgLeuGluAspLeuPheProValLeuAlaTyrValSerValIleSerSerIle 674
QY      688 GCTCTGTAGATGCTTGGACAGTACATATTAACAGTCTATGGCCCTTTCCTACTCATGA 747
Db      674 eSerCysSerLeuSerHisThrIleLysIleProAlaSerPheAlaIleSerSerProPr 694
QY      748 -----CCATCTCTGCTGGAGACTGCTGCTGGTTAAACACCATCATCTGCTAGTTGA 803
Db      694 oThrProHisProGly-----LysAlaAspGlyArgThrHisAsnPheHisLeuPro-L 712
QY      804 TGCCATTATTTCACATCTCCTGGGTTTACACGCTAAAGGGCACCAGCTGCTTAAACA 863
Db      712 euPro-----GlnProAsnThrThrSerArgSerArgProSer---LeuArgg 728
QY      864 GAGAAATAATACATGCTGAGCAGAGGATTAAACAGACTCTCTCACACTACCGGAAGTTTCAG 923
Db      728 lnProSerLeuLeuSerAlaSerAlaMetProThrSerAlaSerThrProAlaSerThrI 748
QY      924 TGCAATTTTCCA 934
Db      748 leAsnProPro 751

RESULT 6
Q8H2R7
ID Q8H2R7 PRELIMINARY; PRT; 374 AA.
AC Q8H2R7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CC chemokine receptor 6.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]_
RP SEQUENCE FROM N.A.
RA Choi Y.K., Fallert B.A., Murphey-Corb M.A., Reinhart T.A.;
RT "Simian Immunodeficiency Virus Dramatically Alters Expression of
RT Homeostatic Chemokines and Dendritic Cell Markers During Infection In
RT Vivo."
RL Blood 0:0-0(2002)
DR EMBL; AF508730; AAN47098.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016493; F:C-C chemokine receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR004067; CC_6_receptor.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PRO1529; CHEMOKINER6.
DR PRINTS; PRO0237; GPCR_RHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 374 AA; 42649 MW; A64A515BABC9E928 CRC64;

```

```

Alignment Scores:
Pred. No.: 1.48 Length: 374
Score: 86.50 Matches: 80
Percent Similarity: 34.89% Conservative: 47
Best Local Similarity: 21.98% Mismatches: 153
Query Match: 2.26% Indels: 84
DB: 6 Gaps: 16

```

US-09-931-733-1 (1-2191) x Q8H2R7 (1-374)

```

QY      37 ATTAATATAAATTTAGTCACTGTAATAATCTAGCTACATAAATTTATATATATATATATATA 96
Db      51 IleAlaTyrSerLeuIleCysValPheGlyLeuGlyAsnIleLeuValValIleThr 70
QY      97 TTCGCACATTTTGTCTTATCAGGAATTTTGTGTATATTTCTACTAATGACATTTTCTGTATG 156
Db      71 PheAlaPheTyrIleLysLysAlaArgSerMetThrAspValTyr---LeuLeuAsnMetAla 89
QY      157 ATAGCAAAATACATTTGTTGGCCACTGCCAAGAGTGGGGGAAAATGCAC--- 208
Db      90 IleAlaAspIleLeuPheVal-LeuThrLeuProPheTrpAlaValSerHisAlaThrGln 109
QY      209 -----ATTTTTAAGCATTTTCTTCTCATCTATGAAAAATTTTCATCTTTGTGGTATGA 261
Db      109 yAlaTrpValPheSerAsnAlaMetCysLysLeuMetLys-----Gln 123
QY      262 ATTATGAGCATTTGCTCGAGTTGCGCCTATAGGTGACCGCAGCATTTGTGTGTATGCTAG 321
Db      123 yIleTyrAlaIleAsnPheAsnCysGlyMet---LeuLeuLeuThrCysIle----- 139
QY      322 TCCCAATTGACACATTTACTGCATATAAAATGATTAATAAGCAAGCAAAAATGTTCT 381
Db      140 -----SerMetAspArgTyrIleAlaIleValGlnAlaThrLysSerPheArgLe 156
QY      382 GAAGGTCTATAACTTCCAAGGGTTCTCTACCTTTTGACCTCTGGATTGGGTTCCCGAGT 441
Db      156 uArgTyrArgThrLeuLeuArgSerLysValIleCysLeuIleValTrpGlyGlySerVa 176
QY      442 AGTGCTTAATATCTCTTCTACACTACTGCTTAAGGCCAACACTCAATGTATCCAGCTAAA 501
Db      176 lVal-----IleSerSerSerThrPheIlePheAsnGlnLysTyrAsnIleGlnGlySe 194
QY      502 GAGTACCTTAAGCCAGTAAAGATTAATCAACAAACACCA-----GAGATGTTTAT 552
Db      194 rAspValCysGluProLysTyrGlnThrValSerGluProIleLysTrpLysLeuLeuMe 214
QY      553 GCCCATATTCCAGATCTCTTCTTCTTCTTCTGACTTACTCTCTTTTGTATGAAGTTATT 612
Db      214 lIeuGlyLeuGluLeuLeuPheGlyPhe-----PheIleProLeuMetValIleIlePh 232
QY      613 CATATGGTTT-----CAAAATCCAGAATTCAGTATTCACATT 648
Db      232 e---TrpTyrMetPheIleValLysThrLeuValGlnAlaGlnAsnSerLysArgHisLy 251
QY      649 CATGGCCCTAGTTTGGATTAAACAAATCAGCTTTCTTTTCTCTCTAGATGCCCTTGAC 708
Db      251 sAlaIleArgValIleIleAlaValLeuValPheLeuAlaCysGlnIlePro----- 269
QY      709 AGTACATATTAAACAGTCTATGGCCCTTCCACTCATGACACATCTCTGCTGCAGATCT 768
Db      270 -----HisMetValLeuLeuValThrAl 278
QY      769 GCTG---CTGGGTAAACACCATCATCTGCTAGTTAGTGAATGCAATTTATTTGACATCTGG 825
Db      278 aAlaAsnLeuGly-AsnMetAsnArgSer-----CysHisSerGluLysLeuLeu- 294
QY      826 GGTTCACACGCTAAAGGGCACCAGTGTCTTAAACAGAAAATACACTGCTGACGA 885
Db      295 -----GlyTyrThrLysThrValThrGluValLeuAlaPheLeuHisC 309
QY      886 GAGGATTAAACAGACTCTCTC----- 905
Db      309 ysCysLeuAsnProValLeuTyrAlaPheIleGlyGlnLysPheArgAsnTyrPheLeuL 329
QY      906 --ACACTACCGGAAGTTTCAGTGCATTTTCCAGATTTTCAAGTCTTTCATGCTGTCTGT 963
Db      329 ysIleMetLysAspLeuTrpCysVal-ArgArgLysTyrLysSerSerGlyPheSerCys 348
QY      964 TGAGAGTCCATATCCACGTTTGTATTATCCAGGAATATATGAGAAAATCTCAGCATTCGT 1023
Db      349 -----AlaGlyArgTyrSerGluAsnIleSerArgGln 359
QY      1024 GAATCT 1029

```

```

Db      360 ThrSer 361
||||
RESULT 7
O35923 ID O35923 PRELIMINARY; PRT; 3343 AA.
AC O35923;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Breast cancer type 2 susceptibility protein.
GN BRCA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=TESTIS;
RX MEDLINE=97384941; PubMed=9242436;
RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA Collins N.K., Fucreal P.A., Bennett L.M., Wiseman R.W.;
RC Cancer Res. 57:3121-3125(1997).
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS. ALSO EXPRESSED
CC IN SPLEEN, SKELETAL MUSCLE, THYMUS, MAMMARY GLAND, HEART, OVARY,
CC PROSTATE, LIVER, LUNG, KIDNEY AND BRAIN.
DR EMBL: U89653; AB71378.1; -.
DR PIR: T42207; T42207.
DR InterPro; IPR002093; BRCA2_repeat.
DR Pfam; PF00634; BRCA2; 8.
DR PROSITE; PS50138; BRCA2_REPEAT; 6.
KW Repeat.
FT DOMAIN 991 2050 8 X 25 AA APPROXIMATE REPEATS.
FT REPEAT 991 1015 BRCL.
FT REPEAT 1204 1228 BRCL.
FT REPEAT 1412 1436 BRCL.
FT REPEAT 1510 1534 BRCL.
FT REPEAT 1645 1669 BRCL.
FT REPEAT 1828 1845 BRCL.
FT REPEAT 1946 1970 BRCL.
FT REPEAT 2026 2050 BRCL.
SQ SEQUENCE 3343 AA; 372213 MW; 653DB110D2302A8D CRC64;

Alignment Scores:
Pred. No.: 2.32 Length: 3343
Score: 86.50 Matches: 49
Percent Similarity: 38.34% Conservatve: 25
Best Local Similarity: 25.39% Mismatches: 57
Query Match: 2.27% Indels: 62
DB: 11 Gaps: 9

US-09-931-733-1 (1-2191) x O35923 (1-3343)
QY 1482 AATGATATATAACCAATAAGTTTTTCTTACAAATTCGATAGCTGAAGAGTTCC 1423
Db |||||
Db 1315 AsnAlaTyrLys-----LeuGluAsnSerAspValSerLysSer 1328
QY 1422 AACATTAGAACTTCTTACCAGCAGACTTACATGTCAGTCATTAGAGACTTATATCCC 1363
Db |||||
Db 1329 ThrSerGlyThr-----ValTyrIleAsn 1336
QY 1362 TTAATGATGCTCAATA-----TAAGAGAAGCAACATGCTTCTTCCCTCCCT 1312
Db |||||
Db 1337 LysGlyAspSerAspLeuProPheAlaGluLysGlyAsnLysTyr-----Pro 1353
QY 1311 TTCCTGTCACAGAGTGGCAGTGCAGAGAGATCTTGTAGGAGCATCTATTGCTACC 1252
Db |||||
Db 1354 GluSerCysThrGlnTyrValArgGluGluAsnAlaGlnIleLysGluSerValSerAsp 1373
QY 1251 AATACTTTACTGCATCTCCTGGTGGGGAATGCT---TCTATGAATATCGCTATTGTT 1195
Db |||||
Db 1374 LeuThrCysLeuGluValMetLysAlaGluGluThrCysHisMetLysSer----- 1390

```

```

QY 1194 TAAATACTACTGAGTATTTCATCCAAATTATTAACTGACAAAGCTGGGCTTTCCAGA 1135
Db |||||
Db 1391 -----SerAspLysGluGlnLeu-Prose 1398
QY 1134 TCAGAAAGTCATTATGAATATGTAAATCCTTCTTAAATTTTATATGAAGAGTTTAACAC 1075
Db |||||
Db 1398 rAspLysMetGluGlnAsn-----MetLysGluPheAsnIle 1410
QY 1074 TAGCTGTCTCTTGACCATGACTAGGTACTTTACCTCATGATTAGATTCAGGAATGCT 1015
Db |||||
Db 1410 eSerPhe-----GlnThrAlaSerGlyLysAsnIle 1420
QY 1014 CAGATTTTCTGCATTATTCTCTGGATATAACAAACGCTGGATAGTAGGACTCTCAACATGACAG 955
Db |||||
Db 1420 eArgValSerLysGluSerLeuAsnLysSerValAsnIleLeuAspGlnGlu----- 1437
QY 954 CATCGAAGACTTGAAAATCTCTGGAAAATCGACTGAAC 918
Db |||||
Db 1438 -ThrGluAspLeuThrValThrSerAspSerLeuAsn 1449

RESULT 8
Q8QGQ5 ID Q8QGQ5 PRELIMINARY; PRT; 409 AA.
AC Q8QGQ5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-carbamoyl-L-amino acid amidohydrolase.
GN LNC.
OS Geobacillus kaustophilus.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1462;
RN [1]
RP SEQUENCE FROM N.A.
RA Chien H.R., Hsu W.;
RT "Cloning, expression, and characterization of the thermostable n-
RT carbamoyl-L-amino acid amidohydrolase from Geobacillus kaustophilus
RT ccc11223."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF425838; AAN31517.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR002933; Peptidase_M20.
DR Pfam; PF01546; Peptidase_M20; 1.
DR PROSITE; PS00037; MYB_1; 1.
KW Hydrolase.
SQ SEQUENCE 409 AA; 44328 MW; 435771D32179A0B1 CRC64;

Alignment Scores:
Pred. No.: 2.27 Length: 409
Score: 85.00 Matches: 37
Percent Similarity: 36.00% Conservatve: 17
Best Local Similarity: 24.67% Mismatches: 40
Query Match: 2.23% Indels: 56
DB: 2 Gaps: 6

US-09-931-733-1 (1-2191) x Q8QGQ5 (1-409)
QY 1509 GTACGAAATTACACTTGTATCTC-----ATAAATGTATATAACACAAATAAGTTTTT 1456
Db |||||
Db 262 ValGlyGlnLeuHisValTyrProGlyGlyIleAsnValIleProGluArgValGluPhe 281
QY 1455 TCTTTACAA-----AATTCGATAGCT 1435
Db |||||
Db 282 ValLeuAspLeuArgAspLeuLysAlaGluValArgAspGlnValTrpAsnAspIleAla 301
QY 1434 GAAGAGATTCACACATTAGAACTTTCTTACCAGCAGACTTACATGTCAGTCATTAGAG 1375
Db |||||
Db 302 SerArgAlaGluThrIleAlaLysGluArgAsnValArgLeuThrThrGluArgLeuGln 321

```



```

QY 1374 ACTTATATCCCTTAATGATGCTCAATATAAGAG-----AAG 1336
Db 322 GluMetAlaProValLeuCysSerGluValValLysGlnAlaAlaGluArgAlaCysLys 341
QY 1335 CAACATGGCTTCTCTCCCTCCCTCTCTC-----1309
Db 342 GlnLeuGlyTyr-----ProProPheTrpLeuProSerGlyAlaAlaHisAspGlyVal 359
QY 1308 -----CTGTGC-----1303
Db 360 GlnLeuAlaProTleCysProIleGlyMetIlePheValArgSerGlnAspGlyValSer 379
QY 1302 -----ACAGAGTGGCAGTGGCAGGAGGAGATCTTGTGTAGGACATCTATGTCTACC 1252
Db 380 HisSerProAlaGlnTrpSerThrLysGluAspCysAlaValGlyAlaGluValLeuTyr 399
QY 1251 AATACTTTACTGCATCTCTCTGGTTGGGGAA 1222
Db 400 HisThrValTrpGlnLeuAlaGlnGlyGlu 409

RESULT 9
Q8PAM4
ID Q8PAM4 PRELIMINARY; PRT; 230 AA.
AC
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein XCC1458.
GN XCC1458.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chamberg F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorxy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AE012246; AAM40755.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 230 AA; 25843 MW; B8CB70C8DFA911E0 CRC64;

Alignment Scores:
Pred. No.: 2.31 Length: 230
Score: 84.50 Matches: 43
Percent Similarity: 39.30% Conservative: 36
Best Local Similarity: 21.33% Mismatches: 76
Query Match: 46 Indels: 46
DB: 16 Gaps: 9

US-09-931-733-1 (1-2191) x Q8PAM4 (1-230)

QY 603 ATCAAAAAGAGTAAGTCACAGAAAGAACAGAGAGTCTGGAATATGCCATAAACATC 544
Db 15 IleArgLysAsnLeuAlaAspHisSerAlaLeuAlaAlaPheArgMetAlaHisGlnIle 34

```

```

QY 543 TGTGGGTGTCGTGATAATAATCTTTTACTGGCTTTAGGGTACTC-----502
Db 35 -----GlyValGlnThrLeuLysAspAlaLysAlaLeu 45
QY 501 -----TTTACGTGGATACATTGAGTGGCCCTTAAGGCAGTAGGTGAAGAGATATTA 448
Db 46 IleGlyPheSerAspLeuHisAsnIleProLeuLysLeuIleGlyHisAlaGly-----63
QY 447 GGCACCTCACTGGCAACCCCAATCCAGAGGT-----CAAAGGGTAGAGAAC 403
Db 64 -----TrpAlaPheAspLeuLysAsnHisValIleHisProLysArgValHisSer 80
QY 402 CCTTGAAGTATTGACCCCTTCAGAACATTTTTCGCTTTTACTTATATCAATTTTATAT 343
Db 81 LysTrpTrpLeu-----IleLeuProPheIleAlaValAlaValLeuIleValLeuAla 98
QY 342 GCAGTAATGCTCAATTTGGGA-----CTAGCATACACAA 307
Db 99 AlaIleLeuGlyAlaIleAlaAlaAsnAspSerLeuValValThrLeuLysGlnThrHis 118
QY 306 ACTGCGGTC---ACCTATAGCGGCAACTCAGACAAATGCTCCATAATTCATACCACAA 250
Db 119 ThrSerIlePheAlaSerGlnThrGluAlaAaGThrGlnThrLeuPheSerValAr3Ser 138
QY 249 GATGAATTTTCATGATGACCAAGAAATGCTTAAATAATGTGCAATTTTCCCCCACTTT 190
Db 139 AspThrLeuThrValGlnGlnCysAsnThrAlaAsnLysGlnValIle-ProProThrPh 158
QY 189 GTGGCAGTGGCCCAACAATTTGTAATTTGCTCATATACAG-----AAAATCA 142
Db 158 eAspProArgAspLeuLysIleLeuCysGlnValTrpGlnAspProAlaLeuLysSerHi 178
QY 141 TTAGTAGAATATACAAAATTTCTGATAAAGCAAAATGTGCAATATATATATATATAT 82
Db 178 s-----LeuAlaLysGluValProLysGlnArgAlaAlaSerLeuLeuGlyPheVa 195
QY 81 A 81
Db 195 I 195

RESULT 10
Q9GTJ5
ID Q9GTJ5 PRELIMINARY; PRT; 507 AA.
AC Q9GTJ5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zinc metalloproteinase 2 ME22 (Fragment).
OS Ancylostoma caninum (Dog hookworm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
OX NCBI_TaxID=29170;
RN [1]
SEQUENCE FROM N.A.
RA Jones B.F., Hotez P.J.;
RT "Cloning and characterization of a zinc-metalloprotease secreted by
RT the invasive stages of Ancylostoma caninum."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF273705; AAG29105.1; -.
DR HSSP; P08473; 1DMT.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008237; F:metalloproteinase activity; IEA.
DR GO; GO:0004245; F:neprilysin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000718; Peptidase_M13.
DR InterPro; IPR008753; Peptidase_M13_N.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01431; Peptidase_M13; 1.
DR PRINTS; PR00786; NEPRILYSIN_N; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.

```

FT NON_TER 1 1
SQ SEQUENCE 507 AA; 58804 MW; 04B4B185D70F6167 CRC64;

Alignment Scores:

Pred. No.: 3,57 Length: 507
Score: 83,50 Matches: 60
Percent Similarity: 37,00% Conservativity: 41
Best Local Similarity: 21,98% Mismatches: 83
Query Match: 2,18% Indels: 89
DB: 5 Gaps: 15

US-09-931-733-1 (1-2191) x Q9H2L2 (1-507)

QY 344 TATATAATGATTAAGTAAAGCCAGAAAATGTTCTGAAGGT-----CATACCTC 397
DB 167 TyrlysAspLeuIleArgSerAlaGlyGlyValMetHisAsnValIleHisAlaPhe 186
QY 398 CAAGGGTCTCTACCTTTCACCTTTCAGCTTGGGTTGGGTTCCAGTGGTCTTAATATCCT 457
DB 187 Gln-----SerMetValAspGlnLeuAspTyrMetSerGlu-----Ala 199
QY 458 TCTTCACCTACTGCTTAAGCCCAACTCAATGATC----- 493
DB 200 ThrlysArgLysAlaIleGluLysSerMetAsnIleIleThrAsnIleAlaPheProAsp 219
QY 494 -----CACGTAAGAGT---ACCTAAAGCCAGTA 520
DB 220 TrpIleMetAspAsnAlaLysLeuAspLeuTyrTyrLysSerIleThrPheAspProThr 239
QY 521 AAGAGTATCAGCAACCCAGACAGATGTTTATGGCCATATTCAGATCCTCTTTCCTTTT 580
DB 240 LysGluAsnTyrTyrAspIleTyrThrLysLeuThrIlePheAsnIleGluAlaGlnTyr 259
QY 581 TCTTGACTTACT-----CCTTTTGTAAAGTATTTCATA--- 616
DB 260 LysHisLeuThrMetAlaThrAlaAspTyrGluGluPheLeuMetProProGlyIleVal 279
QY 617 -----TGTTTCAATCCAGAAATCAGCTATTCACCTATGCGCCCTAGTTTGTGATAAA 670
DB 280 AsnAlaTyrGlnProGluLeuAsnThrIleThrPheProAlaGlyIle----- 296
QY 671 ACAATCAGCTTCTTTTCTGCTGTGAGA-----TGCCCTTGGACA 709
DB 297 -----LeuArgProProTyrPheHisProAspTyrProAlaSer 309
QY 710 GTACATATTAACTCTATGCGCCCTTCCACTCATGACACACATCCCTGCTTGCAGATCTG 769
DB 310 IleLysTyrGlyGlyIleGlyLeuIleAlaGlyHisGlu----- 322
QY 770 CTGCTGGTTAAACACCATCATCTGCTAGTTGATGCAATTTATTGACATCCTGGGTT 829
DB 323 -----LeuIle-HisGlyPheAspValGlnGlyIle 332
QY 830 ACAGCAGCT---AAAGGGCACCAGTGTCTTTAAACAGAGAAAATACATG----- 878
DB 332 LGIntPyrProLysGlyHis---IleSerTyrProLysLysAsnCysIleGlyTyrPm 351
QY 879 -----CTGAGCAGAGATTAAACAGACTCCTCACACTACCGAAGTTTCAGTCAT 928
DB 351 tAspGluGlnSerThrLysGlyPheAsnArgLeu-----AlaGlnCysVa 366
QY 929 TTTCCAGGATTTCAAGCTTCGATGCTGCTGATGTTGAGATGCCATATCCACGTTTGT 988
DB 366 lIleAspGluTyrSerThrPheCysProLeuAspAsnArgThrTyrThrProAsn-CysV 386
QY 989 TA-----TCCAGGAATAATCCAGAAATCTGAGC 1017
DB 386 alaAsnGlyAlaGlnThrGlnGlyGluAsnIleAla 397

RESULT 11

Q9H2L2
ID Q9H2L2 PRELIMINARY; PRT; 333 AA.
AC Q9H2L2;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GPR18-iso.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Dentritic cell;
RA Xu X., Yang Y., Gao G., Xiao H., Chen Z., Han Z.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP261135; ARG44671.1; -. membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00262; G PROTEIN RECEPT F1.2; 1.
SQ SEQUENCE 333 AA; 38034 MW; 06B34C9DFF716415 CRC64;

Alignment Scores:

Pred. No.: 3,75 Length: 333
Score: 83,00 Matches: 29
Percent Similarity: 39,52% Conservativity: 20
Best Local Similarity: 23,39% Mismatches: 47
Query Match: 2,18% Indels: 28
DB: 4 Gaps: 5

US-09-931-733-1 (1-2191) x Q9H2L2 (1-333)

QY 603 ATCAAAAAGAGTAGTCAAGAAAAGCAAGAGAGTCTGGAAATATGCCATAACATC 544
DB 224 LeuLysProLysValLysGluLysSerIleArgIleIleThrLeuLeuValGlnVal 243
QY 543 TCTGGGTTCGCGATAATCTTTTACTGGCTTTAGGGTACTCTTTACGTGGATACATTGA 484
DB 244 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 257
QY 483 GTTGGCTTAAGCAGTAGGTGAAGAGGATATTAAAGGCACCTCACTGGGAACCCCAATCC 424
DB 258 -----MetLeuGlyThrGlyGluAsnSerTyr----- 266
QY 423 AGAGGTCAAGGGTAGAGAACCTTGGAGATATAGCCCTTCAGAACATTTTCTGCCCTT 364
DB 267 -----AsnProTyrGly-----AlaPheThrThrPheLeuMetAsn 278
QY 363 TTA-----CTTATAATCATTTTATATGCAATATGCTCAATTTGGGACTAGCATAC 313
DB 279 LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg 298
QY 312 ACACAACTCGGTACCTTATAGCGGCACTCAGGACAACTCAGGACAACTCCATAATTCATACCAC 253
DB 299 ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg 318
QY 252 AAAGATCAAAATT 241
DB 319 SerGlySerLeu 322
RESULT 12
Q9IB40
ID Q9IB40 PRELIMINARY; PRT; 852 AA.
AC Q9IB40;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C-13 antigen.
GN MAL8P1.57.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;

QY 1532 AA-----GTCAGAGATTTCAGTGTGCACATC-----ACCCGAAAAATGTTAA 1573
 Db luGlnIlePheLysGluSerGluTyrLeuIleTyrPheAsnAsnIleLeuI 662
 QY 1574 CTGTACCCATTAAAGTAATTTCTCATCCGCCATTTCCCTCACCACATGCTCCAGCCCTT 1633
 Db erMetIleLysLysIleLeuPheArgTyrLysIleProGluThrLeuLeuSerAsnPro 682
 QY 1634 GCCAGAACT-----GTTCAATTTCTCTGACTAACAGAA-----AGA 1672
 Db erGluLysAsnSerLeuIleAsnIleLeuHisLeuGlyGluLeuAsnAspPheSerL 702
 QY 1673 AAGCTAAAAAAGTCTGGGAGAGTCTCCACATTTGTTCTTCTTCTTCTTCTTCTTCTGGG 1732
 Db ysGlnCysHisSerLysHisSerLeuMetGlnTyrPheLysIleTyrAsnIleTyrPro 722
 QY 1733 AATAGCAGAATAGGAGCAAGCCA-----G 1756
 Db InAsnGlnAsnGlnSerAlaGluHisLysLeuArgPheAsnAspAspHisLeuI 742
 QY 1757 CACTAGTCTAGCTAAGTACTCAACCAAGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTG 1816
 Db leLysIleSerThrIleHisGlnSerLysGlyLeuGluPhePro----- 756
 QY 1817 ACTTCATTTCTTCTAGCTTTCTGGAGATTACAAACATCTCCGCTTCTGGAAC- 1875
 Db IleThrTyrLeuProPheIleSerCysTyrThrAsnThrLysArgPheSerAsnAsn 776
 QY 1876 -----T 1876
 Db InLysIleLeuGlnSerThrProTyrAlaAsnLeuLeuGlnSerSerThrLeuAsp 796
 QY 1877 TTACTGATTTATCTCCCTCCACACAATAAGCATTTCTTCTTCTTCTTCTTCTTCTTCT 1936
 Db hePheAspLeuGluArgLeuSerGluAspLeuArgLeuLeu-----TyrValAlaIle 814
 QY 1937 CAAGATCTGGACTACTGTTGAAAAAATTTCCAGGTGAGTACTGTTCTTCTTCTTCTTCT 1996
 Db hrArgSerIleTyr-----HisCysSerI 822
 QY 1997 ATGATCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2056
 Db leGlyIleAlaProIleTyrLysSerArgHisLysAsnThrLeuHisThrThrAspL 842
 QY 2057 AGGGTCATGGGATGATGGAGTGGGCTGAGTATTTCAGAAAGTCAAACT-----GAGT 2110
 Db euHisSerAlaLeuGlyTyrLeuIleGlnAsnLysLysProAsnThrProGlnGluL 862
 QY 2111 TGCTTGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2163
 Db euLysGlnCysLeuGlnLysLeuSerIleSerSerAsnLysAsnIleThrPhe 879
 RESULT 14
 Q85483 PRELIMINARY; PRT; 396 AA.
 AC Q85483;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Envelope protein (Fragment).
 OS Rous sarcoma virus.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11886;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Schmidt-Ruppin subgroup A;
 RX MEDLINE=86263387; PubMed=3014723;
 RA Bova C.A., Manfredi J.P., Swanstrom R.;
 RT "env genes of avian retroviruses: nucleotide sequence and molecular
 RT recombinants define host range determinants.";
 RL Virology 152:343-354 (1996).
 DR EMBL; M14901; AAA87245.1; -
 DR PIR; A47611; A47611.

DR GO; 0019031; C:viral envelope; IEA.
 DR InterPro; IPR005166; Avian gp85.
 DR Pfam; PF03708; Avian gp85; 1.
 FT NON_TER 1
 FT CHAIN 57 >396 GLYCOPROTEIN 85.
 FT NON_TER 396 396
 SQ SEQUENCE 396 AA; 43335 MW; 1750744AA1ABA261 CRC64;
 Alignment Scores:
 Pred. No.: 4.46 Length: 396
 Score: 82.50 Matches: 58
 Percent Similarity: 36.15% Conservative: 36
 Best Local Similarity: 22.31% Mismatches: 86
 Query Match: 2.18% Indels: 80
 DB: 15 Gaps: 14
 US-09-931-733-1 (1-2191) x Q85483 (1-396)
 QY 1136 CTGGAAAGCCAGGCTTGTCAAGTTATTATAATTTGGATGATACTTCAGGTAGTATTAA 1195
 Db 61 LeuGluGlnProGly---AsnLeuIrpIleThrTrpAlaAsnArgThrGly----- 76
 QY 1196 ACAATAGCGATTTCATA-----GAAGACATTTTCCCAACACAGGAGATGACGT 1243
 Db 77 --GlnThrAspPheCysLeuSerThrGlnSerAlaThrSerProPheGlnThrCys--- 94
 QY 1244 AAGATTTGTAGCAATAGATGCTCTTACCAAGATCTCTCTGCTGCTGCTGCTGCTG 1303
 Db 95 -----LeuIleGlyIleProSerProIleSerGlu----- 104
 QY 1304 CACAGAAAGGAGGAGGAGAGAGCCATGTTGCTTCTTATATTGAGCATCATTTAAAG 1363
 Db 105 -----GlyAspPheLys 108
 QY 1364 GCAATATAAGTCTTAATGACTACATGTAAGTCTGCTGGTGAAGAAAGTTCTAAAGTTG 1423
 Db 109 Gly-----TyrValSerAspThrAsnCysThrThrLeuGlyThrAsp----- 122
 QY 1424 GAATCTCTTTCAGCTATCGAATTTTGAAGAAAAAATTTATTTGTTGTTATATATATTT 1483
 Db 123 ArgLeuValSerSerAlaAspPheThrGlyGlyProAspAsnSerThrThrLeuThrTyr 142
 QY 1484 ATGAGATACAGTCTTAATTTGTTACATGATATATGATCCATATGTTGGTGAAGTACAG 1543
 Db 143 ArgLysValSerCys-----LeuLeuLeuLysLeuAsnVal 154
 QY 1544 TCAGTGTGCATCATCCCGAAAAATGTAAGTGTACCCATTAAGTAATTTCTCATCCCC 1603
 Db 155 SerMetTrpAspGluProProGluLeu-----GlnLeuLeuGly 167
 QY 1604 CATTTCCCTCACCACATGCTCCAGCCCTTCCAGGAACTGTTTCATTTTCTGACTAA 1663
 Db 168 SerGlnSerLeuProAsnIleThrAsnIleAla----- 178
 QY 1664 CAGAAACGAAAGCTAAAAACACAGTGGGAGGAGTCTCCACATTTGTTCTCT-----ACT 1717
 Db 179 -----GlnIleSerGlyIleThrGlyGlyCysValGlyPheArgProGlnGlyVal 195
 QY 1718 CCATTTTCTTGGGGAATAGCAATAGGAGCAAGC-----CAGCACTGTGCA 1765
 Db 196 ProTrpTyrLeuGlyTyrSerArgGlnGluAlaThrArgPheLeuLeuArgHisProSer 215
 QY 1766 GCTAACTAAGTACTCAACCAAGCCCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1825
 Db 216 PheSerLysSerThrGluPro-----PheThrValThrAlaAspArg-HisAs 232
 QY 1826 TCTTAGGCTTTCTGAGATTAAC---AACATCTGCGGTTCCGTTTCTGGGAACCTTTAC 1880
 Db 232 nLeuPheMetGlySerGluTyrCysGlyAlaTyrGlyTyrArgPheThrAsnMetIyr 251
 RESULT 15
 Q85483
 ID Q85483 PRELIMINARY; PRT; 446 AA.

AC Q8HD85;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 4.
GN ND4.
OS Panulirus japonicus (Japanese spiny lobster).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;
OC Palinuroidea; Palinuridae; Panulirus.
OX NCBI_taxid=6736;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2226630; PubMed=12242015;
RA Yamauchi M.M., Miya M.U., Nishida M.;
RT "Complete mitochondrial DNA sequence of the Japanese spiny lobster,
RL Panulirus japonicus (Crustacea: Decapoda).";
RL Gene 295:89-96(2002).
DR EMBL; AB071201; BAC16324.1; -;
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR001750; Oxidored q1.
DR InterPro; IPR000260; Oxidored q5_N.
DR Pfam; PF00361; oxidored q1; 1.
DR Pfam; PF01059; oxidored_q5_N; 1.
KW Mitochondrion.
SQ SEQUENCE 446 AA; 49684 MW; 4F4FCL1597664EAF CRC64;

Alignment Scores:
Pred. No.: 4.56 Length: 446
Score: 82.50 Matches: 56
Percent Similarity: 37.96% Conservative: 37
Best Local Similarity: 22.86% Mismatches: 77
Query Match: 2.16% Indels: 75
DB: 8 Gaps: 12

US-09-931-733-1 (1-2191) x Q8HD85 (1-446)

QY 173 TTGTTGCCACTGCCAAGAGTGGGG-----AAATGACATTTTAAAGCATTTCTTT 226
Db 11 LeuMetMetPheAlaGlySerTrpGluValGlnPhePheLeuValValSerPhe 30
QY 227 GCTTCATCTATGAAA-----ATTTTCATCTTTTGGTATGAATTATGG 268
Db 31 SerIleSerLeuLysCysAlaHisAspPheMetSerGluLeuGlyTyrMetMetGly 50
QY 269 AGCATTTGCTGATGTCGCCCTATAGTGACCGAGTTTGTGTATGTAGTCCCAAT 328
Db 51 ValAspTyrLeuSerTyrAlaLeuLeuLeuSerAlaTrpValIleAla----- 67
QY 329 TGAGACATTACTGCATATAAATGATTATAAGTAAAGGCGAGAAATGTTCTGAAGGT 388
Db 68 -----LeuMetIleCysSerSerGlnLys---ValMetLysThr 79
QY 389 CATAACTTCCAGGG----- 403
Db 80 TyrAsnPheSerGlyGlyPheIleLeuValAsnMetSerLeuLeuValCysLeuLeu 99
QY 404 ---TTCTTACCTTTCACCTCTGGATTTGGGTTCCAGTGAGTCCCTTAATATCCTTCT 460
Db 100 ThrPheSerSerMetAsnTyrLeuPhePhe-----TyrValSer 112
QY 461 TCACCTACTCCTTAAGGCCCACTCATGTATCCACGTAAGAGTACCCTAAAGCCAGTA 520
Db 113 PheGluSerSerLeuLeuProThrLeuIleLeu-----Ile 124
QY 521 AAAGATTATCACCAGACACCCAGAGATGTTTATGGCC-----ATATTCCAGATC 568
Db 125 LeuGlyTrpGlyTyrGlnProGluArgValGlnAlaGlyLeuTyrMetLeuPheTyrThr 144
QY 569 CTCTTTGCTTTTCTTGACTACTCCTCTTTTGTATAAAGTTATTCATATGTTTCAAATC 628

Db 145 LeuPheAla-----SerLeuProLeuLeuIleCysLeuPheSerLeuTyrGlnAla 161
QY 629 CAGAAATTCAGCTATT-----CACTTCATGCGCCTA 658
Db 162 GlyGlySerLeuIleIleGlyLeuProTyrLysValAspGlnMetAspPheIleSerIle 181
QY 659 GTTTGGATTAAACAATCAGCTTTCTTTCTCTGTAGATGGCTTGGACAGTACATATT 718
Db 182 LeuTrpTyrPheSerSerLeuPheAlaPheLeuValLysLeuPro-----LeuTyrLeu 199
QY 719 AACAGTCTATGGCGCCCTTTCACATCAT-----GACCACATCTCTGCTT 760
Db 200 PheHisIleTrpLeuProLysAlaHisValGluAlaProValAlaGlySerMetIleLeu 219
QY 761 GCAGATCTGCTGCTG 775
Db 220 AlaGlyValLeuLeu 224

Search completed: March 28, 2004, 08:12:06
Job time : 192.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 01:09:44 ; Search time 8590 Seconds
(without alignments)
11055.249 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 2191

Sequence: 1 tacaataatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sv.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2191	100.0	2191	9	S75415S01	S75415 IF116=inter
2	1997.4	91.2	134222	9	AB045363	AB045363 Hcmo sapi
3	1997.4	91.2	300000	9	AP002534	AP002534 Hcmo sapi
4	1995.8	91.1	180742	9	AL359753	AL359753 Human DNA
5	1381.8	63.1	9742	6	AX348912	AX348912 Sequence
6	1330.6	60.7	9742	6	AX348911	AX348911 Sequence
7	274.8	12.5	148081	9	AB045361	AB045361 Hcmo sapi
8	274.8	12.5	179536	2	AC010814	AC010814 Hcmo sapi
9	274.8	12.5	180999	9	AL513205	AL513205 Human DNA
10	240.2	11.0	2775	6	AX188367	AX188367 Sequence
11	240	11.0	2709	9	HUMIFI16A	M63838 Human inter
12	240	11.0	2709	11	G28619	G28619 human STS S
13	240	11.0	4151	9	AF208043	AF208043 Hcmo sapi
14	239.4	10.9	504	6	AX884965	AX884965 Sequence
15	239.4	10.9	504	6	BD024575	BD024575 Sequence
16	238.4	10.9	2716	6	AX281850	AX281850 Sequence
17	233	10.6	2709	9	BC017059	BC017059 Homo sapi
18	231.6	10.6	1093	9	HSU30245	U30245 Human myelo
19	230.2	10.5	640	6	AX884966	AX884966 Sequence
20	230.2	10.5	640	6	BD024576	BD024576 Sequence
21	218	9.9	633	9	HUMZE11G11	AF086545 Homo sapi
22	165.2	7.5	15732	6	AX277933	AX277933 Sequence
23	165.2	7.5	15732	6	AX323620	AX323620 Sequence
24	165	7.5	30614	9	AB045362	AB045362 Homo sapi
25	157.4	7.2	15732	6	AX277932	AX277932 Sequence
26	157.4	7.2	15732	6	AX323619	AX323619 Sequence
27	157	7.2	2070	6	AX833326	AX833326 Sequence
28	157	7.2	2070	9	AK094968	AK094968 Homo sapi
29	150	6.8	31816	9	AL513208	AL513208 Human DNA
30	128.4	5.9	141245	2	AC142815	AC142815 Macaca mu
31	122.2	5.6	1597	9	AY185347	AY185347 Homo sapi
32	122.2	5.6	1624	9	AY185346	AY185346 Homo sapi
33	122.2	5.6	1722	9	AY185345	AY185345 Homo sapi
34	122.2	5.6	1749	9	AY185344	AY185344 Homo sapi
35	120.8	5.5	1668	9	BC020822	BC020822 Homo sapi
36	115.6	5.3	459	6	AX884964	AX884964 Sequence
37	115.6	5.3	459	6	BD024574	BD024574 Sequence
38	106.4	4.9	1892	6	AX477979	AX477979 Sequence
39	99.6	4.5	168981	2	AC126347	AC126347 Homo sapi
40	99.6	4.5	206407	9	AL359266	AL359266 Human DNA
41	98	4.5	147894	2	AC021067	AC021067 Homo sapi
42	97.4	4.4	168900	2	AC024418	AC024418 Homo sapi
43	90.4	4.1	146000	9	AP005433	AP005433 Homo sapi
44	90.4	4.1	180940	2	AP000935	AP000935 Homo sapi
45	90.2	4.1	161047	9	AL772182	AL772182 Human DNA

ALIGNMENTS

RESULT 1
S75415S01
LOCUS S75415S01 2191 bp DNA linear PRI 02-JUN-2000
DEFINITION IF116=interferon-inducible myeloid differentiation transcr:ptional
activator [human, Genomic, 2191 nt, segment 1 of 10].
ACCESSION S75415
VERSION S75415.1 GI:913534
KEYWORDS
SEGMENT 1 of 10
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2191)
AUTHORS Trapani,J.A., Dawson,M., Apostolidis,V.A. and Browne,K.A.

Pred. No. is the number of results predicted by chance to have a

Genomic organization of IFI16, an interferon-inducible gene whose expression is associated with human myeloid cell differentiation: correlation of predicted protein domains with exon organization Immunogenetics 40 (6), 415-424 (1994)

JOURNAL

MEDLINE

PUBMED

REMARK

GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 160709] from the original journal article.

This sequence comes from Fig. 2.

FEATURES

source

1..2191

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

ORIGIN

```

Query Match      100.0%; Score 2191; DB 9; Length 2191;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TACATAATTGAGAGATTTTATTCAGAAATATGCTGATTAATATAAATTTAGTCACTGTA 60
DB 1 TACATAATTGAGAGATTTTATTCAGAAATATGCTGATTAATATAAATTTAGTCACTGTA 60
QY 61 ATAATGTAGTACATAAATTATATATATATATATATATTCGCACATTTTGTCTTTATCAGG 120
DB 61 ATAATGTAGTACATAAATTATATATATATATATATATTCGCACATTTTGTCTTTATCAGG 120
QY 121 AATTTTGTATATTTCTACTAATAGATTTCTGTATGATAGACAAATTAACAATTTGTGC 180
DB 121 AATTTTGTATATTTCTACTAATAGATTTCTGTATGATAGACAAATTAACAATTTGTGC 180
QY 181 CACTGCCCAAGTGGGGAAATGCACATTTTAAAGCATTTCTTGTCTGATCTATGNA 240
DB 181 CACTGCCCAAGTGGGGAAATGCACATTTTAAAGCATTTCTTGTCTGATCTATGNA 240
QY 241 AATTTTCATCTTTTGTGTATGAATTTAGGAGCATTTCTGTAGTTCGGCCCTATAGTGCAC 300
DB 241 AATTTTCATCTTTTGTGTATGAATTTAGGAGCATTTCTGTAGTTCGGCCCTATAGTGCAC 300
QY 301 CGAGTTTGTGTATGCTAGTCCCAATTTGAGACATTTACTGCAATAAATAATATTAAG 360
DB 301 CGAGTTTGTGTATGCTAGTCCCAATTTGAGACATTTACTGCAATAAATAATATTAAG 360
QY 361 TAAAGGCGAGAAAATGTTCTCAAGGGTCATAAATTTCAAGGGTCTCTACCCCTTGACC 420
DB 361 TAAAGGCGAGAAAATGTTCTCAAGGGTCATAAATTTCAAGGGTCTCTACCCCTTGACC 420
QY 421 TCTGATTTGGGTTCCCAAGTGTGCTTAAATCCTTTTCACTTCTGCTTAAAGGCC 480
DB 421 TCTGATTTGGGTTCCCAAGTGTGCTTAAATCCTTTTCACTTCTGCTTAAAGGCC 480
QY 481 AACTCAATGTATCCAGTAAAGAGTACCCCTAAGCCAGTAAAGATTAATACGACACCC 540
DB 481 AACTCAATGTATCCAGTAAAGAGTACCCCTAAGCCAGTAAAGATTAATACGACACCC 540
QY 541 AGAGATGTTTATGGGCATATTCAGATCCTCTTTGCTTTTCTTCACTTACTCTCTTTT 600
DB 541 AGAGATGTTTATGGGCATATTCAGATCCTCTTTGCTTTTCTTCACTTACTCTCTTTT 600
QY 601 GATAAAGTTATTCATATGTTTCAATCCAGAAATTCAGTATTTCACTTCATGGCCCTAGT 660
DB 601 GATAAAGTTATTCATATGTTTCAATCCAGAAATTCAGTATTTCACTTCATGGCCCTAGT 660
QY 661 TTGGATTAAACAATCAGCTTTCTTTTGTCTGTAGATGGCCCTGGACAGTACATATTA 720
DB 661 TTGGATTAAACAATCAGCTTTCTTTTGTCTGTAGATGGCCCTGGACAGTACATATTA 720
QY 721 CAGTCTATGGGCCCTTTTCCACTCATGACCAATCCTGCTTGCAGATCTGCTGGGTTA 780
DB 721 CAGTCTATGGGCCCTTTTCCACTCATGACCAATCCTGCTTGCAGATCTGCTGGGTTA 780
QY 781 ACACCACATCATCTGCTAGTTGATGCCATTTATTTGACATCCTGGGGTTACAGCAGCTA 840

```

```

DB 781 ACACCACATCATCTGCTAGTTGATGCCATTTATTTGACATCCTGGGGTTACAGCAGCTA 840
QY 841 AAGGCGACCCAGTGTCTTAAACAGAGAAAATAACACTGCTGAGCAGAGGATTTAAACAGAC 900
DB 841 AAGGCGACCCAGTGTCTTAAACAGAGAAAATAACACTGCTGAGCAGAGGATTTAAACAGAC 900
QY 901 TCCTCACACTACCGGAAGTTCAGTGCATTTTCCAGGATTTTCAAGTCTTCGATGCTGCA 960
DB 901 TCCTCACACTACCGGAAGTTCAGTGCATTTTCCAGGATTTTCAAGTCTTCGATGCTGCA 960
QY 961 TGTGAGAGTCCACTATCCAGGTTTGTATCCAGGATTAATCGAGAAAATTCGAGCAAT 1020
DB 961 TGTGAGAGTCCACTATCCAGGTTTGTATCCAGGATTAATCGAGAAAATTCGAGCAAT 1020
QY 1021 CGTGAATCTAATCATTTGAGGTAAAGTAACTTAGTCACTGTCAGAGGACAGCTAGTGTTA 1080
DB 1021 CGTGAATCTAATCATTTGAGGTAAAGTAACTTAGTCACTGTCAGAGGACAGCTAGTGTTA 1080
QY 1081 AACTCCTTCATATAAATTTAAGGAAGGATTTTACATATTTCAATAGTATTTTAAACAAA 1140
DB 1081 AACTCCTTCATATAAATTTAAGGAAGGATTTTACATATTTCAATAGTATTTTAAACAAA 1140
QY 1141 AAGCCCGAGCTTGTGAGTATTAATTAATTTGATGAATACCTTTCAGTGTATTTTAAACAAA 1200
DB 1141 AAGCCCGAGCTTGTGAGTATTAATTAATTTGATGAATACCTTTCAGTGTATTTTAAACAAA 1200
QY 1201 TAGCGATTTCTATAGAAGACATTTCCCAACCCAGGAGATTCAGTAAAGTATTTGGTAGCAAT 1260
DB 1201 TAGCGATTTCTATAGAAGACATTTCCCAACCCAGGAGATTCAGTAAAGTATTTGGTAGCAAT 1260
QY 1261 AGATGCTCTTACCAAGATCTCTCTGTCACCTGTCACCTGTCACAGGAAGAGGGGA 1320
DB 1261 AGATGCTCTTACCAAGATCTCTCTGTCACCTGTCACCTGTCACAGGAAGAGGGGA 1320
QY 1321 GAAGAAGCCATTTGTCTTCTTATTTAGAGCATTTTAAAGGAATATTAAGTCTCTAA 1380
DB 1321 GAAGAAGCCATTTGTCTTCTTATTTAGAGCATTTTAAAGGAATATTAAGTCTCTAA 1380
QY 1381 TGACTCACATGTAAGTCTGTGGGTAAAGAGTTCCTAATGTTGGAATCTCTTTTCACTAT 1440
DB 1381 TGACTCACATGTAAGTCTGTGGGTAAAGAGTTCCTAATGTTGGAATCTCTTTTCACTAT 1440
QY 1441 CGAATTTTGAAGAAAATAAATCTTATTTGTTTATATATACATTTATGAGATACAGTGTAA 1500
DB 1441 CGAATTTTGAAGAAAATAAATCTTATTTGTTTATATATACATTTATGAGATACAGTGTAA 1500
QY 1501 TTTCTGTATCATGATATATGCAATTTTGGTGAAGTTCAGAGATTTTCACTGTGCACATCACC 1560
DB 1501 TTTCTGTATCATGATATATGCAATTTTGGTGAAGTTCAGAGATTTTCACTGTGCACATCACC 1560
QY 1561 CGAAAATGTTAATCTGATACCCATTTAAAGTAAATTTCTCATCCCCCATTTCCCTCACCACA 1620
DB 1561 CGAAAATGTTAATCTGATACCCATTTAAAGTAAATTTCTCATCCCCCATTTCCCTCACCACA 1620
QY 1621 TGCTCCAGCCCTTCCAGGAACTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1621 TGCTCCAGCCCTTCCAGGAACTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
QY 1681 AACACTGGTGGGAGGAGTCTCCCATTTTCTTACTTCTTACTTCTTCTGCTGGGAAATGACAG 1740
DB 1681 AACACTGGTGGGAGGAGTCTCCCATTTTCTTACTTCTTACTTCTTCTGCTGGGAAATGACAG 1740
QY 1741 AATAGGAGCAGCCAGCAGTACTAGTACGCTAACTAGTACGCTCAACCAAGGCCCTTTTTCCT 1800
DB 1741 AATAGGAGCAGCCAGCAGTACTAGTACGCTAACTAGTACGCTCAACCAAGGCCCTTTTTCCT 1800
QY 1801 TGTATCTTTGCAGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1860
DB 1801 TGTATCTTTGCAGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1860
QY 1861 TCCGTTCTTGGGAATTTTACTGATTTTATCTCCGCCCTCACAACAATTAAGCATTTGATTCCT 1920

```


Db 99979 TTCTGATCTGGAAGCCAGCGCTTGTGAGTATTAATTAATTTGGATGATTAATCTCAGTAG 100038
QY 1189 TATTTAACAAATAGCGATTTTCATGAGACATTTCCCAACAGGAGATGACGATAAAGT 1248
Db 100039 TATTTAACAAATAGCGATTTTCATGAGACATTTCCCAACAGGAGATGACGATAAAGT 100098
QY 1249 ATTGGTAGCATAGATGCTCTACAGATCTCTCTGCGACATGCCACTCTGTGACAG 1308
Db 100099 ATTGGTAGCATAGATGCTCTACAGATCTCTCTGCGACATGCCACTCTGTGACAG 100158
QY 1309 GAAAGGAGGAG 1368
Db 100159 GAAAGGAGGAG 100218
QY 1369 ATAAGTCTGTAATGACTCACATGTAAGTCTGCTGGGTAAGAAATTTCTAATGTTGGA 1428
Db 100219 ATAAGTCTGTAATGACTCACATGTAAGTCTGCTGGGTAAGAAATTTCTAATGTTGGA 100278
QY 1429 CCTTTCAGCTATCGAATTTTCTGAG-AAAAAAGCTTATTTGTTTATATATACATTTATGA 1487
Db 100279 CCTTTCAGCTATCGAATTTTCTGAGAAAAAAGCTTATTTGTTTATATATACATTTATGA 100338
QY 1488 GATACAAAGTGAATTTCTGTAATGAGATATATGATGATTTGTTGAGTCAAGTCAAGATTTTCA 1547
Db 100339 GATACAAAGTGAATTTCTGTAATGAGATATATGATGATTTGTTGAGTCAAGTCAAGATTTTCA 100398
QY 1548 TGTCACATCACCCGAAATTTGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1607
Db 100399 TGTCACATCACCCGAAATTTGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 100458
QY 1608 TCCCTCACACATGCTCCAGCCCTTCCAGGAACTGTTCAATTTCTGACTAAACAGA 1667
Db 100459 TCCCTCACACATGCTCCAGCCCTTCCAGGAACTGTTCAATTTCTGACTAAACAGA 100518
QY 1668 AACGAAAGCTAAACACATGTTGGGAGGAGTCTCCACATTTCTTCCACTCTCCATTTCTC 1727
Db 100519 AACGAAAGCTAAACACATGTTGGGAGGAGTCTCCACATTTCTTCCACTCTCCATTTCTC 100578
QY 1728 TGGGG-AATGACAGAAATAGGAG 1786
Db 100579 TGGGGCAATAGCAGAAATAGGAG 100638
QY 1787 AGGCTTTTCTTCTGTTTATCTTTGACAGATCTTCAATTTCTAGCGTTCTGAGATTA 1846
Db 100639 AGGCTTTTCTTCTGTTTATCTTTGACAGATCTTCAATTTCTAGCGTTCTGAGATTA 100698
QY 1847 CAACATCTCGGTTCCGTTTCTGGGAATTTTACTGATTTATCTCCGCCCTCACACAAAT 1906
Db 100699 CAACATCTCGGTTCCGTTTCTGGGAATTTTACTGATTTATCTCCGCCCTCACACAAAT 100758
QY 1907 AAGCAATGATCTCGATTTCTGAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTT 1966
Db 100759 AAGCAATGATCTCGATTTCTGAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTT 100818
QY 1967 C---AGTGAATCTGTTCTGATTTGTAATATGATCTGTTCTCTCTGAGT-C 2021
Db 100819 CAGTGAATCTGTTCTGATTTGTAATATGATCTGTTCTCTCTGAGT-C 100878
QY 2022 CCAGAAATCAGAGGGAATCAGTATTT-GTATTCAGGGTCTGATGATGATGATGATGATGATGATGAT 2080
Db 100879 CCAGAAATCAGAGGGAATCAGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 100938
QY 2081 GGCTGAGTATTCAGAAAGTG-AACTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2139
Db 100939 GGCTGAGTATTCAGAAAGTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 100998
QY 2140 GAAGATACAGGCTCTTCTTATT-CACCACAACTGAGGACTGGAACAGAGAAA 2191
Db 100999 GAAGATACAGGCTCTTCTTATTCCACCACAACTGAGGACTGGAACAGAGAAA 101051

LOCUS AP002534 300000 bp DNA linear PRI 17-JUN-2000
DEFINITION Homo sapiens genomic DNA, chromosome 1q22-q23, CDI region, section 3/4.
ACCESSION AP002534
VERSION AP002534.1 GI:8570523
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (sites)
AUTHORS Shina, T., Takishima, N. and Inoko, H.
TITLE DNA sequence analysis of a 1.1-Mb region on chromosome 1q22-q23
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 300000)
AUTHORS Hirakawa, M. and Yamaguchi, H.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2000) Mika Hirakawa, Japan Science and Technology Corporation (JST), Advanced Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081, Japan [E-mail:mika@tokyo.jst.go.jp, URL:http://www-alls.tokyo.jst.go.jp/, Tel:81-3-5214-8491, Fax:81-3-5214-8470]
COMMENT This sequence is conducted by Tokai University as a JST sequencing team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone: +81-463-93-1121, Fax: +81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequence,
Please visit our sequence archive Web site
(http://www-alls.tokyo.jst.go.jp/HGS/top.html)
or send email to webmaster@www-alls.tokyo.jst.go.jp.
FEATURES
source
1.300000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q22-q23"
complement(41083..41209)
/note="PrG-11813/PrG-11814; The location is between each flanking site of PCR primers."
db_xref="GDB:598189"
56196..74273
/gene="MND4"
56196..74273
/gene="MND4"
/note="match MND4 (NM_002432) with account of intervening sequences."
74039..74269
/gene="MND4"
/note="RH95925; The location is between each flanking site of PCR primers."
db_xref="GDB:588364"
74045..74261
/gene="MND4"
/note="RH1742; The location is between each flanking site of PCR primers."
db_xref="GDB:4572245"
181942..182031
/note="SHGC-8115; The location is between each flanking site of PCR primers."
db_xref="GDB:1232045"
complement(226219..226439)
/note="WI-6674; The location is between each flanking site of PCR primers."
db_xref="GDB:588221"
232156..277366
/gene="IFI16"
232156..277366
/gene="IFI16"
misc_feature

/note="match IF116 (NM_005531) with account of intervene
sequences."
STS complement(241289. .241444)
/note="RH48430;The location is between each flanking site
of PCR primers."
STS complement(272376. .272547)
/note="RH34584;The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:355027"
277112. .277257
/gene="IF116"
/note="RH103;The location is between each flanking site of
PCR primers."
STS /db_xref="GDB:4559702"
277120. .277321
/gene="IF116"
/note="RH37786;The location is between each flanking site
of PCR primers."
STS /db_xref="GDB:6455398"
282987. .283075
/note="SHGC-18224;The location is between each flanking
site of PCR primers."
gene /db_xref="GDB:1230561"
complement(284696. .299063)
/gene="AIM2"
misc_feature complement(284696. .299063)
/gene="AIM2"
/note="match AIM2 (NM_004833) with account of intervene
sequences."

ORIGIN

Query Match 91.2%; Score 1997.4; DB 9; Length 300000;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2178; Conservative 0; Mismatches 11; Indels 24; Gaps 15;

QY 1 TACATAATTGAGAGATTTTATTCAGATAATGCTGATTAATTAATAATTTAGTCACTGTA 60
DB 230415 TACATAATTGAGAGATTTTATTCAGATAATGCTGATTAATTAATAATTTAGTCACTGTA 230474
QY 61 ATAATGAGCTACATAA-TTATATATATGATATATATATTCGACA-TTTTGGCTTTATCA 118
DB 230475 ATAATGAGCTACATAA-TTATATATATGATATATATATTCGACA-TTTTGGCTTTATCA 230534
QY 119 GGAATTTTGTATATCTACTATGACTTTCTGATGATAGACACAAATTTACAAATTTGTTG 178
DB 230535 GGAATTTTGTATATCTACTATGACTTTCTGATGATAGACACAAATTTACAAATTTGTTG 230594
QY 179 GCCACTGCCAAGTGGGGGAAATGCAATTTTAAAGCAATTTCTTGGCTTCATCTATG 238
DB 230595 GCCACTGCCAAGTGGGGGAAATGCAATTTTAAAGCAATTTCTTGGCTTCATCTATG 230654
QY 239 AAAATTTCTATCTTTGGTATGAATTTATGAGCAATTTCTGAGTTGCCCTATAGGTG 298
DB 230655 AAAATTTCTATCTTTGGTATGAATTTATGAGCAATTTCTGAGTTGCC-CTTATAGGTG 230713
QY 299 ACCGCAATTTGTGTATGCT-----AGTCCCAATTTGA-GACATTTACTGCATATAAATG 352
DB 230714 ACCGCAATTTGTGTATGCTGGAGTCCCAATTTGAGCAATTTACTGCATATAAATG 230773
QY 353 ATTATAAGTAAAGGAGCAAAATGTTCTGAAGGTTCATAACTTCAAGGGTTCTCTACC 412
DB 230774 ATTATAAGTAAAGGAGCAG-AAAATGTTCTGAAGGTTCATAACTTCAAGGGTTCTCTACC 230832
QY 413 CTTTGAACCTCTGATTTGGGTTCCCAAGTGAAGTACCTTAAAGCCAGTAAAGATTATCAC 532
DB 230833 CTTTGAACCTCTGATTTGGGTTCCCAAGTGAAGTACCTTAAAGCCAGTAAAGATTATCAC 230892
QY 473 TTAAGGCCAACTCAATGTATCCACGTAAAGAGTACCTTAAAGCCAGTAAAGATTATCAC 532
DB 230893 TTAAGGCCAACTCAATGTATCCACGTAAAGAGTACCTTAAAGCCAGTAAAGATTATCAC 230952
QY 533 GAAACCCAGAGATGTTTATGGCCAT-ATTCCAGATCCTCTTTGCTTTTCTTGACTTAC 591
DB 230952 GAAACCCAGAGATGTTTATGGCCAT-ATTCCAGATCCTCTTTGCTTTTCTTGACTTAC 232092

DB 230953 GAAACCCAGAGATGTTTATGGCCATCAATTCAGATCCTCTTTGCTTTTCTTGACTTAC 231012
QY 592 TCCTTTTTCATAAAGTTTATTCATATGTTTTCAAATCCAGAAATTCAGCTATTCACATTCAT 651
DB 231013 TCCTTTTTCATAAAGTTTATTCATATGTTTTCAAATCCAGAAATTCAGCTATTCACATTCAT 231072
QY 652 GGGCCTAGTTTGAATTAACCAATCAGCTTCTTTGCTCTGTAGATGGCTTGGACAGT 711
DB 231073 GGGCCTAGTTTGAATTAACCAATCAGCTTCTTTGCTCTGTAGATGGCTTGGACAGT 231132
QY 712 ACATATTAAACAGTCTATGGG---CCCTTTCCACTCATGACCAATCCTCTGCTTCACATCT 768
DB 231133 ACATATTAAACAGTCTATGGGCCCCCTTTCCACTCATGACCAATCCTCTGCTTCACATCT 231192
QY 769 GCTGCTGGGTTAAACACACATCATCTGCTAGTTGATGCCATTTTATTTGACATCCTCGGGT 828
DB 231193 GCTGCTGGGTTAAACACACATCATCTGCTAGTTGATGCCATTTTATTTGACATCCTCGGGT 231252
QY 829 TACAGCACGCTTAAAGGGCACCCAGTGTCTTAAACAGAAAAATATACACTGCTGAGCAGAG 888
DB 231253 TACAGCACGCTTAAAGGGCACCCACTGTCTTAAACAGAAAAATATACACTGCTGAGCAGAG 231312
QY 889 GATTAAACAGACTCCTTCACTACCGGAGTTCAGTGCATTTTCCAGGATTTTCAAGTCT 948
DB 231313 GATTAAACAGACTCCTTCACTACCGGAGTTCAGTGCATTTTCCAGGATTTTCAAGTCT 231372
QY 949 TCGATGCTGTGTTGAGAGTCCACTATCCAGTTTGTGTTATCCAGGAATAATGAGAA 1008
DB 231373 TCGATGCTGTGTTGAGAGTCCACTATCCAGTTTGTGTTATCCAGGAATAATGAGAA 231432
QY 1009 AATCTCAGCATTCGTGAATCTAATCAATGAGGTAAAGTAACTAGTTCATGCTCAAGGA 1068
DB 231433 AATCTCAGCATTCGTGAATCTAATCAATGAGGTAAAGTAACTAGTTCATGCTCAAGGA 231492
QY 1069 CAGCTAGTGTAAACTCCTTTCATATAAAATTTAAGGAAGATTTTACATATTTCAATAATGACT 1128
DB 231493 CAGCTAGTGTAAACTCCTTTCATATAAAATTTAAGGAAGATTTTACATATTTCAATAATGACT 231552
QY 1129 TTCTGATCTGGAAAGCCAGGCTGTGCTAGTTAATAATTTGGATGAATACTTTCAGTAG 1188
DB 231553 TTCTGATCTGGAAAGCCAGGCTGTGCTAGTTAATAATTTGGATGAATACTTTCAGTAG 231612
QY 1189 TATTAAACAAATAGCGATTTTCATAGACACATTTCCCAACCCAGGAGATGAGTAAAGT 1248
DB 231613 TATTAAACAAATAGCGATTTTCATAGACACATTTCCCAACCCAGGAGATGAGTAAAGT 231672
QY 1249 ATTGTGAGCAATAGATGCTTCTTACCAAGATCTCTCTGCCACTGCCACTCTGTGCACAG 1308
DB 231673 ATTGTGAGCAATAGATGCTTCTTACCAAGATCTCTCTGCCACTGCCACTCTGTGCACAG 231732
QY 1309 GAAAGAGGGGAGAGAACCCATGTTGCTTCTTATAATTTGAGCATCAATTTAAGGAAAT 1368
DB 231733 GAAAGAGGGGAGAGAACCCATGTTGCTTCTTATAATTTGAGCATCAATTTAAGGAAAT 231792
QY 1369 ATAAGTCTCTAATGACTCAATGTAAGTCTGCTGGTAAAGAAAGTTCTAATCTGCAACT 1428
DB 231793 ATAAGTCTCTAATGACTCAATGTAAGTCTGCTGGTAAAGAAAGTTCTAATCTGCAACT 231852
QY 1429 CCTTTGAGTATCGAATTTTGTAAAG-AAAAAACTTATTGTTGTTTATATACATTTATGA 1487
DB 231853 CCTTTGAGTATCGAATTTTGTAAAG-AAAAAACTTATTGTTGTTTATATACATTTATGA 231912
QY 1488 GATAAAGTGTAAATTTGTTTACATGATATATGCCATATTTGTTGAGTCAAGATTTGAG 1547
DB 231913 GATAAAGTGTAAATTTGTTTACATGATATATGCCATATTTGTTGAGTCAAGATTTGAG 231972
QY 1548 TGTGCACATCACCCGAAATAATGTTAACTGTACCCATTAAGTAAATTTCTCATCCCCCAT 1607
DB 231973 TGTGCACATCACCCGAAATAATGTTAACTGTACCCATTAAGTAAATTTCTCATCCCCCAT 232032
QY 1608 TCCCTCTACCATGCTCCAGCCCTTGGCAGGAACTGTTTCAATTTTCTTCTGACTAAACAGA 1667
DB 232033 TCCCTCTACCATGCTCCAGCCCTTGGCAGGAACTGTTTCAATTTTCTTCTGACTAAACAGA 232092

repeat_region /note="ML1A2 repeat: matches 1. .178 of consensus"
21146. .21237
repeat_region /note="46 copies 2 mer tc 78% conserved"
21239. .21441
misc_feature /note="ML1A2 repeat: matches 173. .374 of consensus"
complement(21446. .22253)
repeat_region /note="match: STS: Em:L30626"
22002. .22043
repeat_region /note="14 copies 3 mer gat 85% conserved"
22046. .22292
repeat_region /note="AluSg repeat: matches 38. .284 of consensus"
22379. .22423
repeat_region /note="MER63 repeat: matches 676. .720 of consensus"
22733. .23160
repeat_region /note="L1MEC repeat: matches 1615. .2158 of consensus"
23246. .23873
repeat_region /note="L1R9 repeat: matches 10. .625 of consensus"
24112. .24569
repeat_region /note="L1MCB repeat: matches 834. .1282 of consensus"
24572. .24932
misc_feature /note="L1M4 repeat: matches 3242. .3603 of consensus"
24879. .25247
repeat_region /note="match: GSS: Em:AQ170065"
25175. .25337
repeat_region /note="L1M4 repeat: matches 3693. .3850 of consensus"
25622. .26007
repeat_region /note="MER57B repeat: matches 1. .400 of consensus"
28964. .29026
repeat_region /note="L1P16 repeat: matches 5784. .5840 of consensus"
29030. .29333
misc_feature /note="L1P16 repeat: matches 5844. .6156 of consensus"
complement(29809. .30266)
repeat_region /note="match: GSS: Em:AQ212470"
30008. .30155
repeat_region /note="37 copies 4 mer atat 60% conserved"
31584. .31865
repeat_region /note="AluSq repeat: matches 33. .307 of consensus"
31938. .32109
repeat_region /note="L2 repeat: matches 2356. .2533 of consensus"
32937. .33233
repeat_region /note="L1P4 repeat: matches 5850. .6146 of consensus"
33234. .33693
repeat_region /note="L1P repeat: matches 3649. .4120 of consensus"
34533. .34835
repeat_region /note="AluY repeat: matches 1. .304 of consensus"
complement(35719. .36242)
misc_feature /note="match: GSS: Em:AQ319180"
36382. .36811
repeat_region /note="L2 repeat: matches 2257. .2710 of consensus"
36819. .37123
repeat_region /note="AluX repeat: matches 10. .312 of consensus"
43237. .43414
misc_feature /note="match: GSS: Em:AQ106632"
43416. .43459
repeat_region /note="L1 copies 4 mer taga 100% conserved"
43420. .43578
misc_feature /note="match: GSS: Em:AQ106632"
44428. .44920
repeat_region /note="L1P22 repeat: matches 1. .505 of consensus"
44923. .46258
repeat_region /note="HERVK22 repeat: matches 5503. .6835 of consensus"
46259. .46691
repeat_region /note="L1P3 repeat: matches 5714. .6146 of consensus"
46692. .48529
repeat_region /note="HERVK22 repeat: matches 3655. .5503 of consensus"
48529. .48581
repeat_region /note="HERVK22 repeat: matches 2493. .2545 of consensus"
48574. .48653
repeat_region /note="HERVK22 repeat: matches 2829. .2907 of consensus"
48653. .48796
repeat_region /note="HERVK22 repeat: matches 2327. .2473 of consensus"
49215. .49607
repeat_region /note="HERVK22 repeat: matches 339. .741 of consensus"

repeat_region 49609. .50386
/note="L1P2 repeat: matches 1. .776 of consensus"
50382. .55628
repeat_region /note="L1P2 repeat: matches 900. .6146 of consensus"
55629. .55664
repeat_region /note="18 copies 2 mer aa 86% conserved"
55666. .56016
repeat_region /note="HERVK22 repeat: matches 3. .350 of consensus"
56024. .56512
repeat_region /note="L1R22 repeat: matches 1. .505 of consensus"
56278. .56679
misc_feature /note="match: STS: Em:G16240"
58282. .63993
repeat_region /note="L1P15 repeat: matches -1416. .4402 of consensus"
64119. .64240
repeat_region /note="MIR repeat: matches 100. .232 of consensus"
65689. .65928
repeat_region /note="L1P15 repeat: matches 5920. .6157 of consensus"
67171. .67342
repeat_region /note="L1P3 repeat: matches 5969. .6138 of consensus"
67364. .67417
Query Match 91.1%; Score 1995.8; DB 9; Length 180742;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2177; Conservative 0; Mismatches 12; Indels 24; Gaps 15;
QY 1 TACATAATTGAGAGATTTTATTCAGATAATGCTGATTAATAATAATTTAGTCACTGTA 60
Db 105045 TACATAATTGAGAGATTTTATTCAGATAATGCTGATTAATAATAATTTAGTCACTGTA 105104
QY 61 ATAATGTAGTCAATAA-TTATATATATGATATATATATATGCGACA-TTTTGTGTTTATCA 118
Db 105105 ATAATGTAGTCAATAA-TTATATATATGATATATATATGCGACA-TTTTGTGTTTATCA 105164
QY 119 GGAATTTTGTATTTCTACTAATGACTTTTCTGTATGATAGACAAAATFACAATTTGTTG 178
Db 105165 GGAATTTTGTATTTCTACTAATGACTTTTCTGTATGATAGACAAAATFACAATTTGTTG 105224
QY 179 GCCACTGCCACAAGTGGGGGAAAATGCCAATTTTAAAGCAATTCCTTGTCTTCTCATCTATG 238
Db 105225 GCCACTGCCACAAGTGGGGGAAAATGCCAATTTTAAAGCAATTCCTTGTCTTCTCATCTATG 105284
QY 239 AAAATTTCTATCTTTGTGGTATGAATTTATGAGCAATTTCTGAGTGTCCGCCCTATAGGTG 298
Db 105285 AAAATTTCTATCTTTGTGGTATGAATTTATGAGCAATTTCTGAGTGTCCGCCCTATAGGTG 105343
QY 299 ACCGAGTTTGTGTATGCT-----AGTCCCAATTGA-GACATTAATGATATAAAATG 352
Db 105344 ACCGAGTTTGTGTATGCTGTGAGTCCCAATTTGAGGACATTAATGATATAAAATG 105403
QY 353 ATTATAAGTAAAGGCGAGAAAATGTTCTGAGGGTCATAAATTCGAAAGGTTCTCTACC 412
Db 105404 ATTATAAGTAAAGGCGAG-AAAATGTTCTGAGGGTCATAAATTCGAAAGGTTCTCTACC 105462
QY 413 CTTTGACCTCTGAGTTGGGGTTCCAGTGTGCTTAAATATCTTCTTCTCACCCTAGTGC 472
Db 105463 CTTTGACCTCTGAGTTGGGGTTCCAGTGTGCTTAAATATCTTCTTCTCACCCTAGTGC 105522
QY 473 TTAAGCCCAACTCAATGATCCAGTAAAGAGTACCTTAAAGCCAGTAAAGATTATCAC 532
Db 105523 TTAAGCCCAACTCAATGATCCAGTAAAGAGTACCTTAAAGCCAGTAAAGATTATCAC 105582
QY 533 GAACACCCAGAGATGTTTATGGCCAT-ATTCCAGATCCCTTTTGTCTTCTTCTGACTTAC 591
Db 105583 GAACACCCAGAGATGTTTATGGCCAT-ATTCCAGATCCCTTTTGTCTTCTTCTGACTTAC 105642
QY 592 TCTTTTTTGATAAAGTTTATTCATATGGTTTCAAAATCCAGAAATTCAGCTATTCACTTCAT 651
Db 105643 TCTTTTTTGATAAAGTTTATTCATATGGTTTCAAAATCCAGAAATTCAGCTATTCACTTCAT 105702
QY 652 GGCCTTAGTTTGATTAATAAACAATCAGCTTTCTTTTGTCTCTGTAGATGCCCTTGGACACT 711
Db 105703 GGCCTTAGTTTGATTAATAAACAATCAGCTTTCTTTTGTCTCTGTAGATGCCCTTGGACACT 105762

[illegible]

5164	Db	AAAAAAAAAAAAAAAAAACCAATTAATTA	CTTCTTATATTTAAACATCAATTTAA	AAAAAAT	5101
1369	Qy	ATAAGTCTCTAAATGACTCACATGTA	AGTCTGCTGGGTGAAGAAAGTTCT	TAATGTTGAACT	1428
5104	Db	ATAAATCTCTAATAACTCACATATAA	ATTAATCTACTAAATAAANAATTTCT	ATAATATAA	5045
1429	Qy	CTTTTCAGCTATCGAATTTTGT-AGA	AAAAAACTTATTGTGTGTTTATATA	FACATTTATGA	1487
5044	Db	CTTTTCAACTATCGAATTTTATAAAAA	AAAAAACTTATTATATTTATATACA	TTTATATA	4985
1488	Qy	GATACAAGTGTATTTTCGTTTACATG	ATATATGCCATATTGGTGAAGTCAG	AGATTTCCAG	1547
4984	Db	AATACAAATATAATTTTCGTTTACAA	ATAATATATACCAATATAATAAAT	CAAAATTTCAA	4925
1548	Qy	TGTGCATCACCCGAAAAATGTTAACT	GTACCCATTTAAAGTAATTTCTCATC	CCCCCAT	1607
4924	Db	TATACATCACCCGAAAAATATTAAC	TATACCAATTAATAATTTCTCATC	CCCCCAT	4865
1608	Qy	TCCCTTCACCATGCTCCAGCCCTTGC	CAGGAACTGTTCATTTCTCTGACTA	TCAGAA	1667
4864	Db	TCCCTTCACCATATCTCCAACTTACC	AAAAAACTATTCATTTCTCTAACTA	TCANAA	4805
1668	Qy	AACGAAAGCTAAAAACACTGGTGGG	AGAGTCTCCACATTTGTTCTACTCC	CAATTTCTCT	1727
4804	Db	AACGAAAACTAAAAACACTAATAAAA	AAATCTCCACATATTTCTCTACTCC	CAATTTCTCT	4745
1728	Qy	T-GGGGAATAGCAGAAATAGGAGCAG	CCAGCTAGTACGTAACTAAGTGACTA	CAACA	1786
4744	Db	TAAAACAATAACAAATAAAAAAACA	ACCAACCTAATCAACTAACTAAATA	CTCAACA	4685
1787	Qy	AGGCCCTTTTCTCTGTATCTTTGCA	GAATCTTCATTTTCTTAGGGTTTCT	TGGACATTA	1846
4684	Db	AAACCTTTTTCCTTATATCTTTACA	AAATACCTTCATTTCTTAAGTTTCT	TAABAATTA	4625
1847	Qy	CAACAPCTCGGTTCCGTTTCTGGAA	CTTTACTGATTAATCTCCCCCTCAC	CAAAAT	1906
4624	Db	CAACATCTCAGATTCGGTTTCTAAAA	ACTTTTACTAATTAATCTCCCCCT	CACACAAAT	4565
1907	Qy	AAGCATTTGATTCGTGATTTCTGAG	ATCTCGAATCTGGAATCTGTGTGA	AAAAATTTTC	1966
4564	Db	AAACATTAATTCCTCATTTCTAAAA	ATCTCAAAATCTAAACTATATAA	AAAAAATTTTC	4505
1967	Qy	C---AGGTGAGTACTGTTTCCTCAT	TTTCTGTAATATGATCTTGTCTCT	TCCTTGAAGT-C	2021
4504	Db	CAATAAAATAAATACTATTCTCTAA	TTTTATATAATAATCTTATCTTCT	TCCTTAAAAATCC	4445
2022	Qy	CCAGAATCACAGGGGACAAATCAGTA	TTT-GTTATTCAGGGTCATGGGAT	GATGGGATG	2080
4444	Db	CCAAATATCAAAAAAACAATCAATTA	TTAATTTATTTCAAATCATAAAA	TAATAAAAAATAA	4385
2081	Qy	GGCTGAGTATTCAGAAAAAGT-GAAA	ACTGAGTTGCTTGATATCAATCCT	TCATTTACTTAG	2139
4384	Db	AACTAAATATTCAAAAATAAAAACT	AAATTAATTAATATAAATCTTCAT	TAAATCTTCATTTACTTA	4325
2140	Qy	GAAATACAGGCACTTCTTAAT-CAC	CAACACTGAGGACTGAACAAGAAA		2191
4324	Db	AAAAATAACAACATCTTCTATTCAC	CACTAAAAAATCTTAAACAAAAA		4272

DECLASSIFIED

RESOL 6
AX348911

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS
PORT OF

SOURCE
ORIENTAL

ORGANIZATION

REFERENCES

AUTHORS

TITLE

JOURNAL

AX348911 9742 bp DNA linear PAT 06-FEB-2002
Sequence 369 from Patent WO0202807.

AX348911
AX348911.1 GI:18614946

synthetic construct
synthetic construct
artificial sequences.

¹ Olek, A., Piepenbrock, C. and Berlin, K.
Diagnosis of diseases associated with cell signalling
Patent: WO 0202807-A 369 10-JAN-2002:

FEATURES		Epigenomics AG (DE)		Location/Qualifiers	
source		1..9742		/organism="synthetic construct"	
				/mol_type="unassigned DNA"	
				/db_xref="taxon:32630"	
				/note="Chemically treated genomic DNA (Homo sapiens)"	
ORIGIN					
Query Match		60.7%; Score 1330.6; DB 6; Length 9742;			
Best Local Similarity		79.3%; Pred. No. 6e-259;			
Matches 1755;		Conservative 0; Mismatches 434; Indels 24; Gaps 14;			
QY	1	TACATAATTGAGAGATTTTATTCAGAAATATGCGCTGATTAATATAAATTTAGTCACTGTA	60		
Db	3261	TATATAATTGAGAGATTTTATTTAGAAATATGTTTGAATTAATATAAATTTAGTATTGTA	3320		
QY	61	ATAATGTAGTACAPAA-TTATATATATGATATATATTCGGCACA-TTTTTCCTTTATCA	118		
Db	3321	ATAATGTAGTTATATAAATTTATATATATGATATATATTCGTATATTTTGTGTTTATTA	3380		
QY	119	GGAAATTTTGTATATTTCTACTAATGACATTTCTGATGATAGACACAAATTAACAATTTGTG	178		
Db	3381	GGAAATTTTGTATATTTTATTAATGATTTTTTTGTATGATAGTATAAAATTAATTTGTG	3440		
QY	179	GCCACTGCCACAAAGTGGGGGAAATGACATTTTTTAAAGCAATTTCTTGTCTCACTATG	238		
Db	3441	GTTATTGTTATAAAGTGGGGGAAATGTAATTTTTAAGTATTTTTTGTGTTTATTTATG	3500		
QY	239	AAAAATTCATCTTGTGTTATGAAATATGAGAGCATTTGCTCAGTTTGGCGCTATAGGTTG	298		
Db	3501	AAAAATTTTATTTTGTGTTATGAAATATGAGAGTATTTGTTGAGTTG-TTTTATAGGTTG	3559		
QY	299	ACCGCAGTTTGTGTCT-----ATGCTAGTCCCAATTTGAGACATTTACTGCATATAAAATG	352		
Db	3560	ATCGTAGTTTGTGTGTGTGTGTGTGGAGTTTAAATTTGAGGATATTTATGATATAAATG	3619		
QY	353	ATTATAAGTAAAGCGACAGAAAATGTTCTGAAGGTGCATAACTTCCAGGGTTCTCTACC	412		
Db	3620	ATTATAAGTAAAGGTAG-AAAAATGTTTGAAGGGTTATAATTTTAAAGGGTTTTTATT	3678		
QY	413	CTTTGACCTCTGGATTGGGTTCCAGTGCAGTGCCTTAATATCTCTCTCAGCTACTGCC	472		
Db	3679	TTTTGATTTTGGATTGGGGTTTTTGTGTTGTTTAAATATTTTTTTTATTTATTTATGTT	3738		
QY	473	TTAAGGCCAACTCAATGATCCAGTAAAGAGTACCCCTAAAGCCAGTAAAGATATTCAC	532		
Db	3739	TTAAGGTTAAATTAATGTATTTACGTAAAGAGTATTTTAAAGTTAGTAAAGATTTATAC	3798		
QY	533	GAACACCCAGAGATGTTTATGGCCA-TATTCCAGATCCTCTTCTGCTTTTCTTGACTTAC	591		
Db	3799	GAATATTTAGAGATGTTTATGTTTATTTATTTAGATTTTTTTTGTGTTTTTTTGAATTTAT	3858		
QY	592	TCCTTTTGTGATAAAGTTATTCATATGTTTCCAAATCCAGAAATTCAGCTATTCACITTCAT	651		
Db	3859	TTTTTTTTTGTATAAAGTTATTTATATGCTTTTAAATTTAGAAATTTAGTTATTTATTTAT	3918		
QY	652	GGCCCTAGTTTGGATTAACCAATACGCTTTTGTCTCTGTTAGATGGCCCTTGACAGT	711		
Db	3919	GGTTTTTAGTTTGGATTAATAATTAATTTTGTGTTTTTGTGTTGATGTTGGTTTGGATAGT	3978		
QY	712	ACATATTAACAGTCTATGGG---CCCTTTCCTCACTCATGACACATCTCGCTTGCGAGATCT	768		
Db	3979	ATATATAATAGTTTATGGGGTTTTTTTTTTTATTTAUGATTAATTTTGTGTTGTAGATTT	4038		
QY	769	GCTGCTGGGTTAACCAATACATCGTGTAGTTGATGCCATTTATTTGACATCTCGGGT	828		
Db	4039	GTTGTTGGGTTAATATATATTTATTTGTTAGTTGATGTTATTTATTTTGTATTTTGGGT	4098		
QY	829	TACAGCACGCTAAAGGGCACCGAGTCTCTTAAACAGAGAAAATACACCTGCTGAGCAGAG	888		
Db	4099	TATAGTACGTTAAAGGGTATTTTATTTGTTTTTAAATAGAGAAAATATATTTGTTGAGTAGAG	4158		
QY	889	GATTAAACAGACTCCTCACACTACCGAAGTTTCAGTGCATTTTCCAGGATTTTTCAGTCT	948		
Db	4159	GATTAAATAGATTTTTTATATATATTCGGAAGTTTATGTTATTTTTCAGATTTTTCAGTTT	4218		
QY	949	TCGATGCTGTCATGTTGAGAGTCCACTATCCACGCTTTGTTTATCCAGGAATAATGAGAA	1008		
Db	4219	TCGATGTTGTTATGTTGAGAGTTTATTTATTCAGTTTGTTTATTTAGGAATAATGAGAA	4278		
QY	1009	AATCTGAGCATTCGTGAATCTAATCATTCGAGGTAAAGTAACTTACCTAGTCAATGTCAGAGGA	1068		
Db	4279	AATTTGAGTATTCGTGAATTTAATTTATGAGGTAAAGTAAATTTATGTTATGCTTTAGAGGA	4338		
QY	1069	CAGCTAGTGTAAACTCCTTCATATAAATTTTAAAGGAAGGATTTACATATTTTCATAAAGACT	1128		
Db	4339	TAGTTAGTGTAAATTTTTTTTATATAAATTTTAAAGGAAGGATTTATATAATTTTATAAGTT	4398		
QY	1129	TTCTGATCTGGAAGCCAGCGCTTGTGAGTTTATTAATTAATTTGAGTAAATCTCCAGGTAG	1188		
Db	4399	TTTTGATTTGGAAGTTTAGGTTTGTAGTTTATTAATTAATTTGAGTAAATTTTATGAGTAG	4458		
QY	1189	TATTTAAACAAATAGCGAATTTTCATAGAGACATTTCCCAACCCAGGAGATCGAGTAAAGT	1248		
Db	4459	TATTTAAATAAATAGCGAATTTTATAGTAGATAATTTTTTAAATTTAGAGATGCTGTAAGT	4518		
QY	1249	ATTGCTAGCAATAGATGCTCCTACCAAGATCTCCTCCTGCCACTGCCACTCTGTGCACAG	1308		
Db	4519	ATTGCTAGTAATAGATGCTTTTATTAAGATTTTTTTTGTGTTATTTGTTATTTTGTGATAG	4578		
QY	1309	GAAGGAGGGGAGAGAACCATGTTGCTCTCTTATATTTTGGAGCATCATTTAAGAGGAAT	1368		
Db	4579	GAAGGAGGGGAGAGAAAGTTATGTTGTTTTTTTATATTTTATTTTAAAGGGAAT	4638		
QY	1369	ATAAGTCTCTAATGACTCACATGTAAGTCTGCTGGGTAAAGAAAGTTCTAATGTTGGAAT	1428		
Db	4639	ATAAGTTTTTAAATGATTTATATGTAAGTTTGTGGTAAAGAAAGTTTAAATGTTGGAAT	4698		
QY	1429	CCTTTTCACTATCGAATTTTGTGAAG-AAAAAACCTTATTTGTGTTTATATATACATTTATGA	1487		
Db	4699	TTTTTTTATATCGAATTTTGTGAAGAAAAAAATTTATTTGTGTTTATATATATTTATGA	4758		
QY	1488	GATACAACTGTAATTTTCGTTTACATGGAATATATGCCATATTCGTGGAAGTCAGAGATTTCA	1547		
Db	4759	GATATAAGTGTAAATTCGTTATATGGAATATATGTTATATTTGTTGGAAGTTAGAGATTTTAT	4818		
QY	1548	TGTGCACATACCCGAAAAATGTTAACTGTACCCATTAAGATTAATTTCTCATCCCCCAT	1607		
Db	4819	TGTGTATATTTTCGAAAAATGTTAAATTTGTTATTTTAAAGTAAATTTTTTATTTTATTT	4878		
QY	1608	TCCCTCTCACCATGCTCCAGCCCTTGCAGGAAACTGTTCTATTTTCTCTGACTAACAGA	1667		
Db	4879	TTTTTTTATATATGTTTATGTTTATTTTGTAGGAATTTGTTATTTTTTTTGTATTATAGA	4938		
QY	1668	AACGAAAGCTAAAAACACTGTTGGGAGAGTCTCCACATTTGTTTCTCTACTCCATTTCTC	1727		
Db	4939	AACGAAAGCTAAAAATAATTTGGTGGAGGAGTTTTTATATTTGTTTATTTTATTTTATTTT	4998		
QY	1728	TGGGG-ATATAGCAGATAGGAGCAGCCAGCAGTGTAGTCACTAAGTACTGACTCAACCA	1786		
Db	4999	TGGGGTAAATAGTAATAGGAGTAAAGTTAGTTAGTTAGTTAAATTAAGTGAATTTAATTA	5058		
QY	1787	AGGCTTTTTTCTTGTGTTATCTTTTGCAGATACCTTTCTTACGCGTTTCTCGAGGATTA	1846		
Db	5059	AGGTTTTTTTTTTGTTATTTTGTAGATATTTTATTTTATTTTGTAGCGTTTTTGGAGATTA	5118		
QY	1847	CAACATCTCGGGTCCGTTCTGGGAACTTTACTGATTTATCTCCCGCTCTGTTGAAAAATTTT	1906		
Db	5119	TAAATTTTGGCGTTCCGTTTTTGGGAATTTTATTTGATTTATTTTATTTTATATAAAT	5178		
QY	1907	AAGCATTCATCTCGCTGCTTTCTCGGAACTTTACTGATTTATCTCCCGCTCTGTTGAAAAATTTT	1965		
Db	5179	AAGTATTCATTTTGTATTTTGTGAAGATTTTAAAGATTTTGGATTTATTTGTTGAAAAATTTT	5238		
QY	1966	---CCAGGTGAGTACTGTTCTCTGATTTTGTAAATATGATCTGTTCTCTTCTTCTTGAAG-TC	2021		

Mon Mar 29 09:55:00 2004

```

QY 1123 ATGACATTTCTGATCTGG--AAAGCCAGGCTTCTGAGTTAATAATTAATGATGATACT 1180
Db 136524 ATGACATTTCTGCTCTCTCAAGCTCCAGGTGTATCAGTTAATAATACTGGTGAATA-T 136582
QY 1181 TCAGGTAGTATTAAACAATAGGATTTTCATAGAGACATTTCCCAACAGGAGATGC 1240
Db 136583 GTAGGTACTATTAAACAAA-AAACAATTTCAAGTAGACCCATCCCTATCCAGAGATGA 136641
QY 1241 AGTAAGTATTGGTAGCAATAGATGCTCTACCAAGATCTCTCTGCCACTGCCACTCT 1300
Db 136642 AGTCAAGTCTGGTAGCAATAGATGCTCTACCAAGAT--CTCTGCTCTGCCACTCT 136698
QY 1301 GTGCACAGGAAGAGGAGGAGAA-----GAAGCCATGTTGCTTCTTAATTTGAGC 1353
Db 136699 GTGAACAGGAAGAGGAGGAGAACTTTGGTGTGTTTTCCTTCCCAATTTATTAAGATC 136758
QY 1354 ATCAATTAAGGAATATAAGTCTCTAATGACTCATGTAAGTCTGCTGGTGAAGAAAGT 1413
Db 136759 CTTATTCAAGCGGTACCGGTTTCTAATGACCTATATGTAAGTTAGCTTTATTAAGAAAGT 136818
QY 1414 TCTAATGTTGAACTCTCTTTTCAGCTA 1439
Db 136819 TCTGAGTTGAACTCTCTTCAGCAA 136844

RESULT 10
AX188367 2775 bp DNA linear PAT 06-AUG-2001
LOCUS
DEFINITION Sequence 4062 from Patent WO0142467.
ACCESSION AX188367
VERSION AX188367.1 GI:15139840
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer
Patent: WO 0142467-A 4062 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2775
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 11.0%; Score 240.2; DB 6; Length 2775;
Best Local Similarity 93.3%; Pred. No. 2.2e-38;
Matches 251; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1701 CCACATTTGTTCTTCTACTCCATTTCTCTGGGAATAGCAGATAGAGCAAGCCAGCACT 1760
Db 10 CCGCTGCTACTAGCTGCTGAGAGCCCGGGAATAGCAGATAGAGCAAGCCAGCACT 69
QY 1761 AGTCAGCTAACTAAGTACTCAACCAAGCCCTTTTCTCTGTTATCTTTGAGATACTTT 1820
Db 70 AGTCAGCTAACTAAGTACTCAACCAAGCCCTTTTCTCTGTTATCTTTGAGATACTTT 129
QY 1821 CATTCTTCTAGCGTTCTGAGATTAACAATCTCTGCGTTCGGTTCTGGAACTTTAC 1880
Db 130 CATTCTTCTAGCGTTCTGAGATTAACAATCTCTGCGTTCGGTTCTGGAACTTTAC 189
QY 1881 TGATTTATCTCCCTCTCACAAATAAGCATTTGTTCTGCAATTTCTCAAGATCTCAAG 1940
Db 190 TGATTTATCTCCCTCTCACAAATAAGCATTTGTTCTGCAATTTCTCAAGATCTCAAG 249
QY 1941 ATCTGACTACTGTTGAAAAAATTTCCAG 1969
Db 250 ATCTGACTACTGTTGAAAAAATTTCCAG 278

```

```

RESULT 11
HUMIF116A
LOCUS
DEFINITION Human interferon-gamma induced protein (IFI 16) gene, complete cds.
ACCESSION M63838
VERSION M63838.1 GI:184568
KEYWORDS interferon-gamma inducible protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
AUTHORS Trapani,J.A., Browne,K.A., Dawson,M.J., Rameay,R.G., Eddy,R.L.,
Snow,I.B., White,P.C. and Dupont,B.
TITLE A novel gene constitutively expressed in human lymphoid cells is
inducible with interferon-gamma in myeloid cells
JOURNAL Immunogenetics 36 (6), 369-376 (1992)
MEDLINE 92406263
PUBMED 1526658
COMMENT Original source text: Homo sapiens cDNA to mRNA.
FEATURES
Location/Qualifiers
source
1..2709
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="CTL/NK cell"
1..2709
/gene="IFI 16"
1..264
/gene="IFI 16"
265..2454
/gene="IFI 16"
267..2686
/product="interferon-gamma induced protein"
/protein_id="AAA58683.1"
/db_xref="GI:184569"
/translation="MGKKYKIVLLKGVINDVHFMVKSLLSNDLKMLKREYD
KIQADIMESKFGDAGLGLKIKFIDPTLEDAETLKKELKVGKPALSKRKKKV
HATSPASTSTVKTGEGATKPAQKQKSTKEKAGKPKGSKVSEOTQPPSPAGAGMS
TAMGRSPSTJUSAPENSSSTENPKTVAKQVTPRNVLQKRPVIVKVLSTPKPEY
ETPEMEKIMFATVATOTFFHVKVLNTSLKEKFNKKIILISDYLEDSLELVNBE
STVSEAGNQTFFVFNKIINRAKELKILDKHQASGNIIVGVFMHLKKTNNQKITY
EIODRGKMDVVGTCQCHNIPCBEGLKQLFCFLRKNQMSKLSISMHSIFQIKKKT
NPRNDPKMKILPOEQQLPVPEASTTTPESHRLTPQMPPTSPSSFTTSPSSFTKSDTIS
KXNDPMQKILKESHPFPMTSIGAESHPHTPQMPPTSPSSFTTSPSSFTKSDTIS
EVSIEDSAQSDLKAEVWLVNATESFVYEPKEQKMFATVATENEVPR/KVFNIDLE
KFTPKIATANYVCRNGELEVYPTLVDVADNRNMEIPKGLRSASVTPKINQLCS
QTKGSFVNGVFEVHKGNVRGEFTYEIQDNTGRMEVVVHGRLNTINCEEGDKLUTSF
ELAPKSGNTGELRSVHSHIKVIKTRKNKKDILNPDSSMETSPTDFP"
2677..2686
/gene="IFI 16"
polyA_signal
ORIGIN
Query Match 11.0%; Score 240; DB 9; Length 2709;
Best Local Similarity 100.0%; Pred. No. 2.4e-38;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1730 GGAATAGCAGAAATAGGAGCAAGCCAGCACTAGTCAGTCACTAAGTCACTCAACCAAGG 1789
Db 1 GGAATAGCAGAAATAGGAGCAAGCCAGCACTAGTCAGTCACTAAGTCACTCAACCAAGG 60
QY 1790 CTTTTTCTTCTGTTATCTTTTCAGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAA 1849
Db 61 CTTTTTCTTCTGTTATCTTTTCAGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAA 120
QY 1850 CATCTCGCGTTCCTGTTCTGGAACTTTACTGATTTATCTCCCTCTCACAAATAAG 1909
Db 121 CATCTCGCGTTCCTGTTCTGGAACTTTACTGATTTATCTCCCTCTCACAAATAAG 180
QY 1910 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCAG 1969
Db 181 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCAG 240

```



```
|||||
Db 121 CATCTGGGTTCCGTTCTTGGAACTTACTGATTTATCTCCCCCTCACAAATAAG 180
QY 1910 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCAG 1969
Db 181 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCAG 240

RESULT 14
AX884965
LOCUS AX884965 504 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 828 from Patent EP1033401.
ACCESSION AX884965
VERSION AX884965.1 GI:40040653
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
Expressed sequence tags and encoded human proteins
Patent: EP 1033401-A 828 06-SEP-2000;
Genset (FR)
FEATURES
source Location/Qualifiers
1..504
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
264..>503
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE99079.1"
/db_xref="GI:40040654"
/translation="MSVVMGKKYKNIVLLKGLVINDYHFRMVKSLLSNDLKLKMR
EYDKIQIADIMEEKPRGDAGLKLKIFEDIPLE"
ORIGIN
Query Match 10.9%; Score 239.4; DB 6; Length 504;
Best Local Similarity 99.6%; Pred. No. 4.7e-38;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1729 GGGGAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 1788
Db 11 GGGCAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 70
QY 1789 GCCTTTTTCCTTGTATCTTGCAGATCTTCAATTTCTTAGCGTTTCTGGAGATTACA 1848
Db 71 GCCTTTTTCCTTGTATCTTGCAGATCTTCAATTTCTTAGCGTTTCTGGAGATTACA 130
QY 1849 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 1908
Db 131 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 190
QY 1909 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCA 1968
Db 191 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCA 250
QY 1969 G 1969
Db 251 G 251

RESULT 15
BD024575
LOCUS BD024575 504 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD024575
VERSION BD024575.1 GI:22565798
KEYWORDS JP 2001269182-A/821.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 504)
Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
Sequence tag and encoded human protein
Patent: JP 2001269182-A 821 02-OCT-2001;
GENSET
OS Homo sapiens (human)
PN JP 2001269182-A/821
PD 02-OCT-2001
PR 24-FEB-2000 JP 2000118773
PI JEAN BAPTISTE DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
PI JORDAN
PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
C12N5/10,
PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC Key Location/Qualifiers
FH CDS 264..503.
FT CDS
FEATURES
source Location/Qualifiers
1..504
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 10.9%; Score 239.4; DB 6; Length 504;
Best Local Similarity 99.6%; Pred. No. 4.7e-38;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1729 GGGGAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 1788
Db 11 GGGCAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 70
QY 1789 GCCTTTTTCCTTGTATCTTGCAGATCTTCAATTTCTTAGCGTTTCTGGAGATTACA 1848
Db 71 GCCTTTTTCCTTGTATCTTGCAGATCTTCAATTTCTTAGCGTTTCTGGAGATTACA 130
QY 1849 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 1908
Db 131 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 190
QY 1909 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCA 1968
Db 191 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAAATTTCCA 250
QY 1969 G 1969
Db 251 G 251
Search completed: March 28, 2004, 05:37:50
Job time : 8599 secs
```